Geochemistry Geophysics Geosystems G3

Published by AGU and the Geochemical Society

Article
Volume 13, Number 5
2 June 2012
Q05Z43, doi:10.1029/2012GC004046
ISSN: 1525-2027

Towards the robust selection of Thellier-type paleointensity data: The influence of experimental noise

Greig A. Paterson
Paleomagnetism and Geochronology Laboratory, Key Laboratory of Earth's Deep Interior, Institute of Geology and Geophysics, Chinese Academy of Sciences, Beijing 10029, China
(greig.paterson@mail.iggcas.ac.cn)

Andrew J. Biggin
Geomagnetism Laboratory, Geology and Geophysics, School of Environmental Sciences, University of Liverpool, Oliver Lodge Laboratories, Oxford Road, Liverpool L69 7ZE, UK

Yuhji Yamamoto
Center for Advanced Marine Core Research, Kochi University, Kochi 783-8502, Japan

Yongxin Pan
Paleomagnetism and Geochronology Laboratory, Key Laboratory of Earth's Deep Interior, Institute of Geology and Geophysics, Chinese Academy of Sciences, Beijing 10029, China

[1] The process of data selection in paleointensity studies is an essential step to ensure data fidelity. There is, however, no consensus as to the best approach to consistently select data with most studies using arbitrarily defined thresholds for selection. We present a new numerical model that simulates the variability of paleointensity data from hypothetical ideal samples acquiring a thermoremanent magnetization (TRM) by incorporating experimental noise, which has been constrained using over 75,000 data measurements. Using Monte Carlo analyses, we investigate the behavior of simulated data and characterize the distributions of parameters typically used to select paleointensity data. We use the 95th percentiles of the distributions to define thresholds for the maximum likely parameter values that can result from experimental noise. These represent values below which we cannot distinguish non-ideal behavior from noise. We find that a number of parameters are highly sensitive to noise and laboratory field strength (e.g., partial TRM, pTRM, check *CDRAT* and pTRM tail check δt^*); this sensitivity may diminish their ability to identify non-ideal behavior. The fractional (f) dependence of some parameters and the proportion of inaccurate results provide justification for $f \geq 0.35$ when selecting data from both Thellier-Thellier and Coe protocol experiments. The manifestation of noise in the original Thellier method, however, is different to that of methods that use zero-field heating steps. This suggests that the data selection procedure for the Thellier method should be different, but it also suggests that, contrary to previous analyses, the accuracy and scatter of results from this method are more sensitive to noise than methods that use zero-field heating steps. The general approach taken here is shown to be a powerful means of understanding the behavior of selection parameters and has the potential to be extended to models incorporating non-ideal behavior resulting from alteration and multidomain grains.

Components: 3300 words, 9 figures, 4 tables.

Keywords: error analysis; experimental noise; non-ideal behavior; paleointensity.

Index Terms: 1521 Geomagnetism and Paleomagnetism: Paleointensity; 1594 Geomagnetism and Paleomagnetism: Instruments and techniques; 1599 Geomagnetism and Paleomagnetism: General or miscellaneous.

Received 12 January 2012; **Revised** 23 April 2012; **Accepted** 23 April 2012; **Published** 2 June 2012.

Paterson, G. A., A. J. Biggin, Y. Yamamoto, and Y. Pan (2012), Towards the robust selection of Thellier-type paleointensity data: The influence of experimental noise, *Geochem. Geophys. Geosyst.*, *13*, Q05Z43, doi:10.1029/2012GC004046.

Theme: Magnetism From Atomic to Planetary Scales: Physical Principles and Interdisciplinary Applications in Geosciences and Planetary Sciences

1. Introduction

[2] Developing long-term records of geomagnetic field variation is fundamental to understanding geodynamo evolution. Determining the absolute paleointensity of the ancient geomagnetic field, however, is difficult and time consuming, and experiments are prone to high failure rates. This has been the motivation behind an increase in efforts to understand and detect the causes of paleointensity failure [e.g., *Krása et al.*, 2003; *Leonhardt et al.*, 2004; *Biggin et al.*, 2007; *Draeger et al.*, 2006; *Yu and Tauxe*, 2006; *Fabian*, 2009].

[3] During a paleointensity experiment a samples natural remanent magnetization (NRM) is progressively replaced by a laboratory thermoremanent magnetization (TRM), which is acquired in a magnetic field of known strength. The ratio of the NRM-to-TRM gives us an estimate of the ratio of the ancient-to-laboratory field strength. Causes of paleointensity failure are broadly termed as "non-ideal" behavior and may include, but are not limited to, NRM that is not of thermal origin, magnetomineralogical alteration (in nature or during laboratory heating), the presence of large magnetic grains within a sample, non-linear TRM acquisition, or anisotropy of TRM. As our understanding of non-ideal behavior has progressed, new checks and selection criteria have been developed to identify and reject non-ideal behavior [e.g., *Riisager and Riisager*, 2001; *Krása et al.*, 2003; *Paterson*, 2011]. Defining appropriate thresholds for these checks to facilitate the detection and exclusion of non-ideal behavior is a key aim in paleointensity studies, but one that has yet to be fully realized.

[4] Numerical models are an increasingly used tool to characterize and understand paleointensity data [e.g., *Fabian*, 2001; *Leonhardt et al.*, 2004; *Biggin*, 2006; *Fabian*, 2009], but existing models do not predict the high degree of variability that is seen in real paleointensity data. In this study we have developed a phenomenological paleointensity model to investigate the variability of data obtained from hypothetical ideal single domain (SD) samples as a result of random noise being introduced during the various stages of the paleointensity experiment. The incorporation of experimental noise into the results from ideal samples allows us to define lower threshold values for parameters designed to identify non-ideal behavior (i.e., threshold values below which non-ideal behavior cannot be distinguished from background noise). This, in turn, allows us to place the selection of paleointensity data on a less arbitrary footing, which makes the selection process more effective and ensures that the results of paleointensity studies are more reliable.

[5] In section 2, we give detailed descriptions of the sources and magnitudes of experimental noise that may affect a paleointensity experiment and outline the data used to constrain this. Paleointensity theory and the approach used to model data from ideal samples are outlined in section 3, as well as how experimental noise is incorporated into the model. The effects of experimental noise on two samples with contrasting demagnetization behavior are investigated in section 4 and in section 5 we describe how experimental noise influences various parameters used to select paleointensity data.

2. The Sources and Magnitudes of Experimental Noise

[6] There are a number of potential sources of experimental noise that can influence a paleointensity experiment (e.g., temperature uncertainties, measurement uncertainties). The main sources of experimental noise incorporated into this model are described below and summarized in Table 1. Given that noise is a statistical phenomenon it is necessary to define the probability density function (PDF) of a particular noise source and, where possible, this has been achieved using experimental data. For many

Table 1. Description of the Sources of Experimental Noise Incorporated into the Model[a]

Name	Description	Physical Magnitude	Model Magnitude
δT_{Repeat}	The variation in the reproducibility of heating temperature.	0.14°C	$\sim 2.4 \times 10^{-4}$ ($=\frac{0.14}{580}$)
δT_{Grad}	Temperature variation due to thermal gradients within the furnace.	0.08°C	$\sim 1.4 \times 10^{-4}$ ($=\frac{0.08}{580}$)
δT_{HT}	Temperature variation that results from variation in hold time. Modeled assuming an average hold time of 1600 s with a standard deviation of 30 s.	Varies as a function of heating temperature.	See §2.1.3.
δT_{CR}	Temperature variation that results from variation in cooling rate. Both average cooling time and standard deviation of cooling time vary with heating temperature.	Varies as a function of heating temperature.	See §2.1.4.
δT_{Total}	Total temperature uncertainty.	Varies as a function of heating temperature.	$\sqrt{\delta T_{Repeat}^2 + \delta T_{Grad}^2 + \delta T_{HT}^2 + \delta T_{CR}^2}$
δF_{Lab}	Variations in the applied laboratory field.	~2–4 nT	$0.000075 \times F_{Lab}$
δF_{Res}	Residual field variations.	~20 nT	$0.0005 \times F_{Lab}$
$\delta\phi$	Angular uncertainty of the applied field.	Estimated to be the same as for $\delta\theta$.	Same as for $\delta\theta$.
$\delta M_{(x,y,z)}$	Measurement uncertainty of the magnetometer.	0.36% of the respective *x*-, *y*-, or *z*-axis.	$0.0036 \times$ **RM**
δBG	Background noise of the magnetometer.	~96% of measurements have $\delta BG < 0.02\%$ of the NRM.	Cauchy distributed, see §2.3.2.
$\delta\theta$	Angular uncertainty introduced through reorientation of the sample between measurements.	Weibull distributed with a median of ~1.4°.	Weibull distributed, see §2.3.3.

[a]Unless stated otherwise, all uncertainties are described as standard deviations of a normal distribution. With the exception of $\delta\phi$, which is estimated, all uncertainties are based on experimental data or technical specifications.

sources of experimental noise normality can be assumed (details given below) and the PDFs can be described as standard deviations about a mean value. Where estimates of the standard deviation are made from experimental data, values have been corrected for sample size [e.g., *Holtzman*, 1950].

2.1. Temperature Uncertainties

2.1.1. Repeat Heating

[7] A paleointensity experiment involves multiple heatings to the same temperature and the reproducibility of the repeat heatings can have an effect on the data obtained, particularly over a temperature range where remanence is rapidly lost or gained. Detailed temperature measurements during heating and cooling of the Natsuhara-Giken TDS-1 thermal demagnetizers at the Center for Advanced Marine Core Research, Kochi University, Japan, were performed during routine Thellier-type and Shaw-type paleointensity measurements. Temperatures were measured using the inbuilt thermocouple located in the sample region of the furnace and were taken every 10 s throughout the entire heating/cooling cycle. An example of a full thermal cycle and the quantification of the data are given in the auxiliary material.[1] Data were collected from a total of 58 heatings to various set temperatures. Typically three or less repeat heatings were performed for most set temperatures, but 29 heatings to a set temperature of ~610°C were measured. The standard deviation of the peak temperature from these 29 heatings is 0.18°C; however, there is some indication that the hold time has an influence. Fifteen data have a hold time of ~2500 s and 14 have a hold time of ~3500 s, with standard deviations of 0.14°C and 0.22°C, respectively. This indicates that longer hold times lead to greater deviations in the repeatability of peak temperature, which may be due to increased likelihood of larger fluctuations in the peak temperature with increasing hold time. The model uses an average hold time of 40 minutes (2400 s), therefore a temperature reproducibility standard deviation (δT_{Repeat}) of ~0.14°C is taken to be a reasonable estimate. In our model we simulate the effects of experimental noise by randomly drawing an effective peak temperature from a continuous PDF. The Kolmogorov-Smirnov (KS) test cannot reject the null hypothesis that variations in peak temperature reproducibility are normally distributed about their mean value at the 0.05 significance level. We therefore assume that the effective peak temperatures are normally distributed.

[1]Auxiliary materials are available in the HTML. doi:10.1029/2012gc004046.

2.1.2. Furnace Thermal Gradients

[8] Within paleointensity furnaces it is common for thermal gradients to exist along the length of the furnace. In the presence of such gradients, error in the repositioning of a sample within the furnace may contribute to variations in the reproducibility of the peak temperature during heating. Measurements of the peak temperature as a function of position within an ASC TD48 thermal demagnetizer at the Institute of Geology and Geophysics, Chinese Academy of Sciences (IGGCAS) indicate peak thermal gradients of ~0.32°C/cm. These gradients are approximately independent of the set temperature (details are given in the auxiliary material). No data are available to constrain the repositioning of a sample within a furnace. For simplicity, however, we assume that sample position is normally distributed about the initial position and that ~95% of the time the sample can be repositioned to within 0.5 cm of the initial position (i.e., a standard deviation of 0.25 cm). For a fully loaded paleointensity furnace, where the spacing between samples is small, this is a reasonable estimate. For a partially loaded oven, however, this may be an underestimate. Combined with the measured thermal gradients, this suggests that the variation in peak temperature due to thermal gradients (δT_{Grad}) is on the order of ~0.08°C, is independent of the set temperature (T), and follows a normal distribution of the form $\mathcal{N}(T, \delta T_{Grad}^2)$. This treatment of δT_{Grad} ignores any thermal gradients that may exist over the sample and may result in inhomogeneous heating of the samples.

2.1.3. Furnace Hold Time

[9] The acquisition and demagnetization of TRM is controlled by a time-temperature relation such that remanence acquired or demagnetized at high temperature for a short period of time can be equivalent to remanence acquired or demagnetized at a lower temperature, but for a longer period of time [*Néel*, 1949; *Pullaiah et al.*, 1975]. This thermoviscous behavior is an inherent property of magnetic materials and during laboratory experiments that involve thermal activation these effects can become important. During a paleointensity experiment, a sample is typically held at temperature for a period time, which allows the sample to reach an equilibrium temperature with the furnace. If, however, the sample is held at temperature for a longer period of time it is possible that excess remanence will be demagnetized or gained relative to previous heating steps. Therefore, the effective temperature of (un)

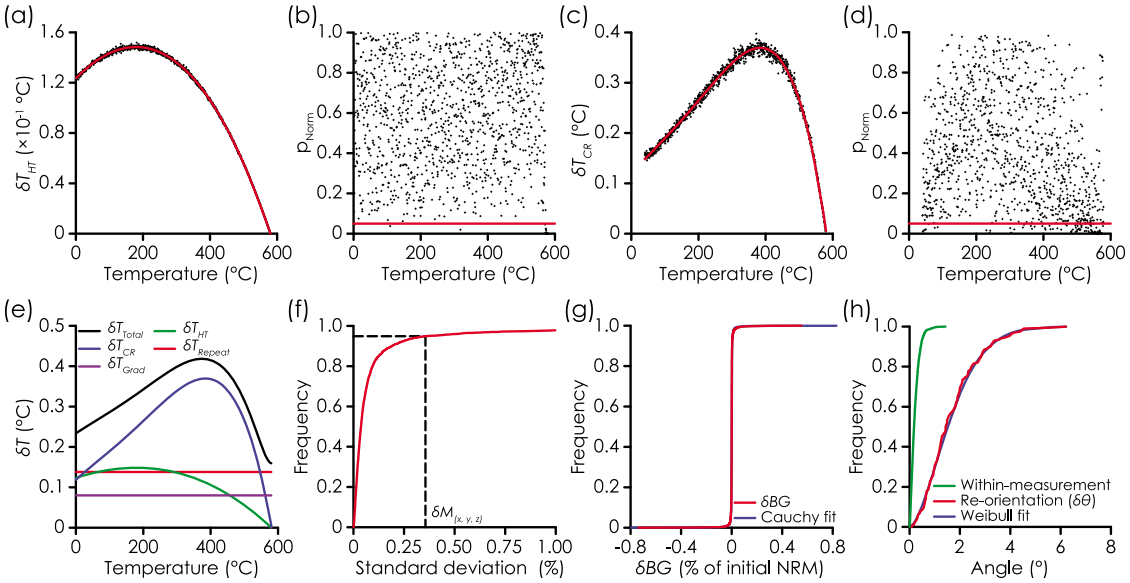

Figure 1. (a) Standard deviation of temperature that results from variations in hold time. Black dots are determined from the Monte Carlo method described in section 2.1.3 and the red line is the best-fit polynomial. (b) Kolmogorov-Smirnov (KS) test probabilities that temperature uncertainties that result from hold time variations are normally distributed. The red line indicates the 0.05 significance level below which the assumption of normality can be rejected. (c) Standard deviation of temperature that results from variations in cooling rate. The symbols are the same as in Figure 1a. (d) KS test probabilities that temperature uncertainties that result from cooling rate variation are normally distributed. The red line is the same as in Figure 1b. (e) Total temperature deviation (δT_{Total}) and individual contributions as function of the set temperature. (f) Empirical cumulative distribution function (ECDF) of the standard deviation of magnetometer measurements as a percentage of the respective NRM vector component. The dashed line marks the 95th percentile, which was used to define $\delta M_{(x,y,z)}$. (g) ECDF of measured magnetometer background noise (red line) and the best-fit Cauchy distribution (blue line). (h) ECDFs of angular deviation from the within-measurement noise (green line), reorientation uncertainty ($\delta\theta$; red line), and the best-fit Weibull distribution to the reorientation distribution (blue line).

blocking is controlled by the set temperature and the hold time.

[10] In the model an average hold time of 2400 s is used and based on the thermal demagnetizer temperature data (with a hold time of ~2500 s) the KS cannot reject the null hypothesis (0.05 significance level) that the hold time data are normally distributed with a mean of 2500 s and standard deviation of 30 s. We therefore adopt a standard deviation of hold time of 30 s. This hold time refers the period of time during which the thermal demagnetizer remains at peak temperature and may not reflect the time that the sample is at peak temperature. First-order lumped capacitance thermodynamic calculations [see, e.g., *Incropera et al.*, 2007] based on the thermal demagnetizer data indicate that a standard 2.5 cm paleomagnetic sample will reach ≳99% of the peak temperature after ~800 s of hold time. Therefore the effective hold time is ~1600 ± 30 s and this value is used for the calculation of temperature variations due to hold time variations.

[11] Assuming that SD magnetite is the magnetic carrier, we can use the simple time-temperature relation outlined by *Pullaiah et al.* [1975] to translate the deviation about the average hold time into a deviation about the average effective temperature (equivalent to deviations about the set temperature, T). For magnetite, *Pullaiah et al.* [1975] defined the time-temperature relation of magnetization by:

$$\frac{T_1 \ln(f_0 t_1)}{M_s^2(T_1)} = \frac{T_2 \ln(f_0 t_2)}{M_s^2(T_2)}, \quad (1)$$

where f_0 is the attempt frequency of thermal fluctuations (≈10^{10} Hz), $M_s(T)$ is the saturation magnetization at temperature T, and t_1, T_1, t_2, T_2 are the equivalent time-temperature pairs. For $M_s(T)$ variations we assume that $M_s(T) \propto (1 - T)^{0.39}$ (details are given in the Auxiliary Material) [see also *Tauxe*, 2010].

[12] Given an average effective hold time (t_1 = 1600 s), an average set temperature ($T_1 = T$), and a random hold time (t_2) drawn from a normal distribution with mean t_1 and a standard deviation of 30 s, an effective demagnetization temperature (T_2) can be numerically calculated from equation (1). For each set temperature this procedure is repeated 10^4 times to define the distribution of effective temperatures and the standard deviation (δT_{HT}) of this distribution about T_1 is calculated. The variation of δT_{HT} as a function of set temperature is shown in Figure 1a and the probabilities that the effective temperatures are normally distributed (determined by the KS test) are shown in Figure 1b.

[13] The deviation of effective temperatures varies as a function of the set temperature and can be approximated by a cubic polynomial of the form $\delta T_{HT} = aT^3 + bT^2 + cT + d$, where $a = -2.764 \times 10^{-10}$, $b = -6.679 \times 10^{-7}$, $c = 2.678 \times 10^{-4}$, and $d = 0.123$. With the exception of few points, the KS test cannot reject the null hypothesis that the effective temperatures that result from variation in hold time are normally distributed at the 0.05 significance level. In order to maximize the efficiency of the model, this simple empirical approximation is adopted. When the set temperature is zero (i.e., no heating) or when $T \geq T_c$ (the Curie temperature), δT_{HT} is set to zero.

2.1.4. Cooling Rate

[14] The time-temperature dependence of TRM also manifests as a dependence on the rate of cooling during remanence acquisition/demagnetization [*Dodson and McClelland-Brown*, 1980; *Halgedahl et al.*, 1980] and variations of cooling rate between heatings to the same temperature will contribute to experimental noise. The thermal demagnetizer data indicate that the time taken to cool to ambient temperature is exponentially related to the set temperature, as would be expected from Newtonian cooling. It is also found that the cooling time and the standard deviation of cooling time approximately follow a linear relation in log-log space. These two relations (outlined in the auxiliary material) allow the cooling time and standard deviation of cooling time to be estimated as a function of the set temperature.

[15] For SD magnetite, *Dodson and McClelland-Brown* [1980] described the effective demagnetization temperature (T_2) by:

$$T_2 \simeq \frac{\epsilon_0 M_s^r(T_2)}{k_B \ln(f_0 t_h)}, \tag{2}$$

where $\epsilon_0 M_s^r(T_2)$ is the energy barrier to thermal fluctuations, ϵ_0 is the energy barrier at $T = T_0 = 0$, r is an anisotropy dependent constant (for magnetite dominated by shape anisotropy $r \approx 2$), k_B is the Boltzmann constant, and t_h is the demagnetization timescale (t_h = 1600 s). The initial energy barrier, ϵ_0, can be described by:

$$\frac{\epsilon_0 M_s^r(T_1)}{k_B T_1} = \ln\left(\frac{\gamma f_0}{\dot{T}_1}\right) - \ln\left(\frac{\epsilon_0 Q(T_1)}{k_B T_1}\right), \tag{3}$$

where

$$Q(T_1) = r M_s^{(r-1)}(T_1)\frac{dM_s(T_1)}{dT_1} - \frac{M_s^r(T_1)}{T_1}. \tag{4}$$

In equation (3), $\ln(\gamma)$ is Euler's constants and $\dot{T}$ is the cooling rate. In the framework of *Dodson and McClelland-Brown* [1980], T_1 and T_2 are the "natural" and "laboratory" (un)blocking temperatures, respectively. In the context of our model T_1 can be viewed as the set (un)blocking temperature and T_2 the effective (un)blocking temperature that is experienced by the sample. As before, $M_s(T)$ variations are approximated by $M_s(T) \propto (1 - T)^{0.39}$.

[16] Given a mean cooling time (t_1) from the set temperature (T_1), an approximate cooling rate ($\dot{T}_1 = \frac{(T_1 - T_0)}{t_1}$) can be calculated (based on the cooling time relation derived from the thermal demagnetizer data). The mean effective demagnetization temperature (T_2) is then calculated by numerically solving equation (2). Variations in the cooling rate, $\dot{T}_1$, will results in variations in T_2 and, for a fixed T_1, a distribution of $\dot{T}_1$ will give rise to a distribution of T_2 values.

[17] The cooling rate of a repeat heating to T_1 is calculated by randomly drawing a cooling time (t_2) from a normal distribution with mean t_1 and standard deviation determined from the empirical relation obtained from the thermal demagnetizer data. Equation (2) is numerically solved for T_2 and the process is repeated for each set temperature 10^3 times to build the distribution of T_2.

[18] The temperature deviation that results from variations in cooling rate (δT_{CR}) is a function of the set temperature and can be approximated by a fourth order polynomial of the form $\delta T_{CR} = aT^4 + bT^3 + cT^2 + dT + e$, where $a = -1.701 \times 10^{-11}$, $b = 9.487 \times 10^{-9}$, $c = -1.457 \times 10^{-6}$, $d = 7.777 \times 10^{-4}$, and $e = 0.118$ (Figure 1c). As is the case for δT_{HT}, for most points the KS test cannot reject the null hypothesis that the effective temperatures are

normally distributed at the 0.05 significance level (Figure 1d). This assumption, however, breaks down as $T \rightarrow T_c$ and δT_{CR} drops rapidly to zero. In our model this should only affect a small number of points. We therefore assume normality and use the above empirical approximation to model the temperature errors due to cooling rate variation. A key assumption of our treatment of cooling rate is that the NRM and TRM are blocked over the same timescale (before the consideration of noise), which negates the need to apply a cooling rate correction to the final paleointensity estimate.

2.1.5. Total Temperature Variation

[19] Given that normality and independence can be assumed for the four above described sources of temperature noise, the total variance of temperature can be expressed as the sum of the squares of the repeat heating, thermal gradient, the hold time, and cooling rate variances:

$$\delta T_{Total}^2 = \delta T_{Repeat}^2 + \delta T_{Grad}^2 + \delta T_{HT}^2 + \delta T_{CR}^2. \quad (5)$$

The contributions that each of these make to the total temperature uncertainty are shown in Figure 1e. Temperature fluctuations in a paleointensity experiment are modeled by randomly drawing an effective temperature from a normal distribution of the form $\mathcal{N}(T, \delta T_{Total}^2)$.

2.2. Field Uncertainties

2.2.1. Applied Laboratory Field

[20] The variation of the applied field (F_{Lab}) during a paleointensity experiment can be assessed by considering the stability and reproducibility of the applied current. The Yokogawa Electric Corporation model 7651 constant current power supply that is used at Kochi University for the applied fields, outputs currents of ~30–50 mA to generate fields of ~30–50 μT in the Natsuhara-Giken furnace (typical fields used during a paleointensity experiment). The repeatability of the applied field between steps can be estimated by the resolution of the applied current, which limits the accuracy of reproducing the same current. For this power supply the applied current resolution is ±100 nA, which corresponds to an applied field reproducibility of ±0.1 nT. The temporal stability of the applied current may also influence F_{Lab} during a paleointensity experiment. The stability of the applied current is ±(0.0015% + 0.3 μA) over a 24 h period. For typical fields and for a typical duration of applied field (~2 h), this corresponds to temporal stability of ~0.06–0.09 nT.

[21] As is the case for thermal gradients small field gradients exist within a paleointensity furnace. Measurement of the variation of applied field as a function of sample position within the ASC TD-48SC furnace at the University of Southampton and the Pyrox paleointensity furnace at the IGGCAS indicate that peak gradients in the applied field are linearly correlated with the set applied field strength (shown in the auxiliary material). This linear relationship corresponds to constant field gradients that are ~0.03% of the applied field. This is on the order of 9–15 nT/cm for applied fields of 30–50 μT. If we consider the treatment of sample position given in section 2.1.2, the corresponding standard deviation of field gradients is 2.25–3.75 nT, or ~0.0075% of the applied field. This treatment of sample repositioning deals only with the variation in the magnitude of the applied field and variations of the direction of the applied field with respect the samples are discussed below.

[22] The influence of temporal stability and applied current reproducibility are at least an order of magnitude small than the effects of field gradients and are not considered in the model. We therefore estimate the applied field uncertainty (δF_{Lab}) as $0.000075 \times F_{Lab}$. Values of the effective applied field are randomly drawn from a normal distribution of the form $\mathcal{N}(F_{Lab}, \delta F_{Lab}^2)$.

2.2.2. Residual Fields

[23] Residual fields within thermal demagnetizers are variably reported to be on the order of ~5 nT to <150 nT [e.g., *Yu and Dunlop*, 2002; *Yamamoto et al.*, 2003; *Pan et al.*, 2005; *Yamamoto and Tsunakawa*, 2005; *Draeger et al.*, 2006; *Yu and Tauxe*, 2006; *Biggin et al.*, 2007; *Shcherbakova et al.*, 2008; *Böhnel et al.*, 2009; *Paterson et al.*, 2010a; *Zheng et al.*, 2010]. Excluding poorly constrained estimates (i.e., those that only report residual fields less than a peak value) the average of the reported residual fields is ~24 nT with a standard deviation of ~20 nT, which corresponds to ~0.04–0.06% of typically applied fields used during TRM acquisition. In the model the residual field intensity deviation (δF_{Res}) is taken to be 0.05% of F_{Lab}. Most of the reported residual field estimates only note the total intensity and not direction. Data from *Zheng et al.* [2010] suggest that the orientation of the residual fields is variable between ovens, but no data from the same oven between heating steps are available for this study. In the model δF_{Res}

only describes the residual field intensity. We assume that the residual fields are oriented along the same axis as the applied field and orientation noise is added to the total applied field (see below). Residual field values in the model are randomly drawn from a normal distribution of the form $\mathcal{N}(0, \delta F_{Res}^2)$ (i.e., on average, residual fields are zero).

2.2.3. Field Orientation

[24] Uncertainties in the orientation of the applied or residual field with respect to the sample are also considered. These may arise from fluctuations in the direction of the applied field due to residual fields or from differences in sample orientation between heating steps. No data are available to constrain the magnitude of this misorientation uncertainty ($\delta\phi$). In the model we incorporate field orientation uncertainties using the same procedure as is used for incorporating uncertainties due to sample misorientation during measurement, which is described in section 2.3.3.

2.3. Measurement Uncertainties

2.3.1. Magnetometer Measurement Noise

[25] Magnetometer measurements of a magnetic remanence vector are subject to random noise, which can be determined by taking multiple measurements of a remanence vector. Automated alternating field (AF) demagnetization was carried out on 64 volcanic samples using the 2G Enterprise 760 SQUID Magnetometer at the IGGCAS. At each demagnetization step the NRM vector was measured 5 times on each axis to determine the axis mean and associated measurement standard deviation. A total of 3615 measurements of the NRM vector components were obtained. In general, the noise is proportional to the intensity of the respective axis measurement with ~95% of all the data having a measurement standard deviation ≤0.36% of the mean axis measurement (Figure 1f). In the model measurement uncertainty ($\delta M_{(x,y,z)}$) is taken as 0.0036 × **RM**, where **RM** is the remanence vector (e.g., the NRM vector). Measurement noise is incorporated into the model by adding remanence fluctuations drawn from $\mathcal{N}(0, \delta M_{(x,y,z)}^2)$ to the individual x, y, and z components of the "measured" remanence vector.

2.3.2. Background Noise

[26] A total of 9,858 magnetometer background measurements obtained during the automated measurement of 151 samples were used to estimate background noise of the IGGCAS 2G magnetometer (Figure 1g). Background noise (δBG) is typically low with ~96% of all measurements being less than 0.02% of the initial NRM. The distribution of δBG can be approximated by a Cauchy distribution with location parameter $a = 1.785 \times 10^{-4}$ and shape parameter $b = 8.729 \times 10^{-4}$ and values are randomly drawn from this distribution. Although the measured $|\delta BG|$ is ≤0.8%, there is a finite probability that the best fit Cauchy distribution will produce unrealistic levels of background noise. Therefore, $|\delta BG|$ is set to a maximum of 0.8% of the initial NRM. Testing of the model and the final results, however, indicates that δBG makes a negligible contribution to the final remanence data (see section 4). Magnetometer drift was found to be of a comparable magnitude or less than background noise and has not been considered in this model.

2.3.3. Sample Reorientation

[27] Manual handling of a sample will lead to uncertainties in the reorientation of the sample with respect to the measurement axes of the magnetometer. Eight volcanic samples were measured 20 times on the IGGCAS 2G magnetometer without being removed from the sample tray. The samples were half standard size (1.1 cm length, 2.5 cm diameter) and plastic holders were used to fix the samples in place. The angular difference between the 20 measured directions and the mean direction of the 20 measurements for each sample was calculated. This is the within-measurement angular deviation that results from measurement noise (i.e., $\delta M_{(x,y,z)}$). The samples were then removed from and replaced back onto the sample tray and measured 20 times. This process was repeated a total of 25 times. For each sample the angles between the 25 repeat mean directions and the mean of those 25 directions were calculated. The angles from all samples were combined to estimate the distribution of angles that results from removing and replacing a sample into the magnetometer ($\delta\theta$; Figure 1h).

[28] The distribution of within-measurement angular deviation tends to lower values compared with $\delta\theta$, which suggests that $\delta M_{(x,y,z)}$ was sufficiently averaged and that $\delta M_{(x,y,z)}$ and $\delta\theta$ can be treated as independent sources of noise. The distribution of $\delta\theta$ in radians can be well approximated by a Weibull distribution with scale parameter $a = 0.033$ and shape parameter $b = 1.633$. If the vectors are transformed onto a two-dimensional plane where angular deviation can be given a sense of rotation (i.e., positive or negative rotation) the distribution

of $\delta\theta$ would correspond to a normal distribution with a standard deviation of $\sim 2^\circ$ (0.034 radians, i.e., $\approx a$). For full sized standard paleomagnetic samples it may be expected that $\delta\theta$ should be lower. For mini-samples (1 cm length, 1 cm diameter), which are now commonly used for paleointensity experiments, $\delta\theta$ is likely to be of a similar magnitude. Angular noise is incorporated by randomly drawing a value for $\delta\theta$ from the above Weibull distribution and rotating the remanence vector by this angle around a randomly generated rotation axis. It may be expected that misorientation is most likely to be due to preferential rotation of the sample about the z-axis (i.e., the axis of the orientation arrow). The measured misorientation, however, includes sample translation and quantifies the total misorientation as an effective angle of rotation. The result is that the axes of rotation are randomly distributed on a unit sphere with no preference for rotation about a single axis (see the auxiliary material for further details). The same procedure and angular distribution is used for $\delta\phi$. In this case the orientation vector of the effective applied field (applied and/or residual fields) is rotated.

3. Modeling an Ideal Paleointensity Experiment

3.1. Paleointensity Protocols

[29] A number of basic paleointensity methods exist (e.g., Shaw-type or Thellier-type). In this study we have modeled the most commonly used Thellier-type methods: the original Thellier-Thellier [*Thellier and Thellier*, 1959], the Coe [*Coe*, 1967], the Aitken/Walton [*Aitken et al.*, 1988; *Walton*, 1979], and the IZZI [*Yu et al.*, 2004] protocols. These procedures are based on the principle that the NRM of a paleomagnetic sample is progressively replaced by a laboratory TRM acquired in a know field (F_{Lab}). The strength of the ancient geomagnetic field (F_{Anc}) can be determined from:

$$\frac{F_{Anc}}{F_{Lab}} = \frac{NRM_{Anc}}{TRM_{Lab}}. \quad (6)$$

[30] When the NRM is progressively replaced by a laboratory TRM, multiple estimates of this ratio can be determined, which provides a more robust estimate of F_{Anc}. Analysis is typically performed on an Arai diagram [*Nagata et al.*, 1963], which plots the NRM remaining after demagnetization against the TRM gained after remagnetization to the same temperature. The best-fit slope through selected points on the plot provides the estimate of $\frac{F_{Anc}}{F_{Lab}}$.

[31] The protocols differ in the order in which the sample is de-/remagnetized. The Thellier-Thellier protocol involves a first heating step in an applied field, F_{Lab}. The sample is then reheated to the same temperature and cooled in a field of the same strength, but with opposite polarity ($-F_{Lab}$). The vector sum of these two resultant magnetizations is twice the NRM remaining after heating and the vector difference is twice the TRM gained. In the Coe protocol the first heating step occurs in zero-field, which allows the NRM remaining to be directly measured. The second heating is in an applied field. The Aitken protocol reverses this sequence with the first step being the in-field step. The IZZI protocol alternates between the Aitken sequence (in-field, zero-field; IZ) and the Coe sequence (zero-field, in-field; ZI).

[32] In our modeled experiments we use 14 temperature steps between ambient temperature and the Curie temperature (T_c). The model assumes SD magnetite is main magnetic carrier and the experiment uses temperature steps of 0, 75, 150, 225, 300, 375, 450, 500, 530, 560, 565, 570, 575, and 580°C. Where appropriate, the procedure incorporates both partial TRM (pTRM) checks and pTRM tails checks, which are standard tests for non-ideal behavior. These checks were conducted at alternating temperature steps (i.e., pTRM tail checks at 75, 225, 375, etc. and pTRM checks at 75, 150, 300, etc.). The pTRM checks cover a continuous range, such that the check to 75°C was performed after a peak temperature of 150°C, the check to 150°C was performed after a peak temperature of 300°C, and so on. Although pTRM checks are routinely performed, pTRM tail checks are not and the effects of omitting these from the experimental procedure will be discussed in section 6.4.

[33] For a hypothetical ideal sample in the absence of noise, all Thellier-type protocols will yield identical results. With the exception of the Thellier-Thellier protocol all protocol yield identical results when subject to experimental noise. For brevity, in sections 4 and 5 we present results from the most widely used protocol, the Coe protocol, but the influence of experimental protocol and the differences seen from the Thellier-Thellier protocol are discussed in section 6.4.

3.2. Defining a Blocking Function

[34] An ideal non-interacting SD sample obeys Thellier's laws of independence, additivity and reciprocity [*Thellier*, 1938; *Thellier and Thellier*, 1959]. The law of independence states that pTRMs

imparted over different, non-overlapping temperature intervals are completely independent in direction and intensity. Additivity is the property that the sum of all pTRMs acquired between T_c and ambient temperature should be equal to the total TRM acquired by cooling from T_c to ambient temperature in a single step. Thellier's law of reciprocity states that a pTRM acquired over a particular temperature interval, say pTRM(T_2, T_1), is completely removed by reheating to T_2 in zero field. This assumption is equivalent to saying that the blocking temperature and the unblocking temperature are identical.

[35] Given these properties the blocking and unblocking of remanence carried by ideal non-interacting SD samples can be described by identical distributions of (un)blocking temperatures. In the case of a paleointensity experiment the NRM remaining after demagnetization to temperature T_i and the TRM acquired after remagnetization to T_i can be phenomenologically described by:

$$NRM(T_i) = F_{Anc}\int_{T_i}^{T_c} f(T)dT, \tag{7}$$

$$TRM(T_i) = F_{Lab}\int_{0}^{T_i} f(T)dT, \tag{8}$$

where $f(T)$ is the distribution of (un)blocking temperatures. Practically speaking, in a Coe protocol paleointensity experiment the TRM gained cannot be directly measured. The total magnetization (J), which is the summation of the TRM gained and NRM remaining, is measured:

$$\begin{aligned} J(T_i) &= NRM(T_i) + TRM(T_i) \\ &= F_{Anc}\int_{T_i}^{T_c} f(T)dT + F_{Lab}\int_{0}^{T_i} f(T)dT. \end{aligned} \tag{9}$$

The TRM gained is determined from:

$$TRM(T_i) = J(T_i) - NRM(T_i). \tag{10}$$

These equations can be generalized to the case of a three-dimensional remanence vector by describing the fields and respective remanences in terms of x, y, and z vector components and assuming that the blocking function, $f(T)$, is independent of field orientation (i.e., the common assumption that paleomagnetic samples are isotropic).

[36] The function $f(T)$ should be such that $\int_0^{T_c} f(T)\,dT = 1$ (i.e., all blocking occurs below T_c). Several functions have been proposed to model $f(T)$ [e.g., *Kono and Tanaka*, 1984; *Fabian*, 2001], in this model, however, a beta distribution is used. This distribution is preferred for a number of reasons. First, unlike other proposed distributions a beta distribution exists only in the range [0, 1], which constrains all blocking to occur between ambient temperature and T_c. Second, NRM thermal demagnetization data from 2115 published volcanic samples [*Tauxe et al.*, 2004a, 2004b; *Huang et al.*, 2005, 2006, 2007; *Zhu et al.*, 2008; *Liu and Zhu*, 2009; *Pan et al.*, 2005; *Paterson*, 2009; *Tauxe and Kodama*, 2009; *Paterson et al.*, 2010b; *Muxworthy et al.*, 2011; *Qin et al.*, 2011] and 102 unpublished samples (basalts from the Emeishan Large Igneous Province, SW China) were used to assess the quality of fit provided by various functions. Assuming a beta distribution of unblocking temperatures provides the best overall fit to the real data when compared with other tested distributions (details are given in the auxiliary material). We use the best-fit beta distributions to the real data as input to our simulations. Since the beta distribution is bound to [0, 1], all temperatures in the model have been normalized by the T_c of magnetite (580°C).

3.3. Incorporating Noise Into the Model

[37] In their model of paleointensity noise, *Kono and Tanaka* [1984] assumed individual sources of remanence variance to be independent and normally distributed and used Gaussian error propagation to describe the variances of NRM and TRM. It should be noted, however, that while the base variations may be assumed to follow a normal distribution, when transformed into remanence variations normality may not be preserved. Consider temperature variations in the situation where the NRM is demagnetized to the Curie temperature. If the effective temperature is less than T_c by, for example, 2°C the sample is under-demagnetized, if the temperature is identically T_c then the sample is fully demagnetized, and in the case where the temperature exceeds T_c by 2°C the sample is also fully demagnetized. Although the temperature variation is symmetric, due to the fact that T_c limits the unblocking range, the remanence uncertainty is asymmetric and may not be approximated by normality. The extent of the asymmetry depends on a number of factors such as the (un)blocking spectrum of the sample. It may also be masked by variations that are normally distributed in remanence space (e.g., measurement uncertainties). To overcome this, the present model incorporates variations at a level where the distribution is known or can be reasonably assumed and numerically propagates these variations

(through equations (7)–(10)) to determine the remanence variations.

[38] Experimental noise is added in a sequence that represents the physical procedure of a paleointensity experiment. Temperature noise (from all sources) and effective applied field noise (applied and residual fields, and field orientation noise) are added into equations (7) and (9). This represents the "heating" phase of an experiment. Following this, measurement, measurement orientation, and background noise are added to the "measured" NRM and J vectors. The noise is then numerically propagated into the TRM vector through equation (10). The model uses continuous integration of equations (7) and (9) as opposed to discrete maps of (un)blocking temperatures [e.g., *Fabian*, 2001; *Biggin*, 2006]. As a consequence, the model holds no "memory" of previous treatments and this has to be explicitly incorporated. For example, the NRM remaining after demagnetization to T_i can be described by:

$$NRM(T_i) = F_{Anc}\int_{T_i}^{T_c} f(T)dT + F_{res}\int_{0}^{T_i} f(T)dT. \quad (11)$$

If, for example, the remagnetization step heats to $T_i' > T_i$, strictly adhering to equation (9) neglects the excess NRM demagnetized. The remagnetization step must therefore be explicitly described by:

$$J(T_i') = NRM(T_i) + (F_{Lab} + F_{res}')\int_{0}^{T_i'} f(T)dT - F_{Anc}\int_{T_i}^{T_i'} f(T)dT - F_{res}\int_{0}^{T_i} f(T)dT, \quad (12)$$

where the last two terms represent the demagnetization of excess NRM and residual field magnetizations, respectively.

4. The Distribution of Uncertainties at the Sample Level

[39] Before investigating the effects of experimental noise on paleointensity selection parameters we first examine the influence of noise on the data obtained from two samples: One with a broad (un)blocking temperature range (Sample 1), the other with narrow (un)blocking close to T_c (Sample 2). The blocking functions of these samples are based on the best-fit functions to real demagnetization data and were chosen due to their contrasting (un)blocking behavior.

[40] The modeled experiment uses the Coe protocol with both pTRM and pTRM tail checks at the temperature steps outlined in section 3. The laboratory field is identical in strength and direction to the NRM acquisition field. Idealized (i.e., without noise) NRM unblocking and TRM blocking for Sample 1 and Sample 2 are shown in Figures 2a and 2c, respectively. Idealized Arai plots are shown in Figures 2b and 2d, respectively. Since the propagation of errors is achieved numerically, a Monte Carlo approach with 10^4 simulations was used to determine the underlying distribution from which the remanence data are drawn. With this approach noise is randomly drawn from the distributions outlined in section 2, added to each step of the experiment and numerically propagated through the calculations of the TRM and the check differences. The error bars in Figures 2b and 2d indicate the range of NRM and TRM values obtained from the Monte Carlo simulations.

4.1. NRM and TRM Distributions

[41] At most temperature steps and for both samples the KS test cannot reject the hypothesis that the NRM and TRM are normally distributed about their respective means at the 0.05 significance level (Figures 2e and 2g). For both samples the assumption of normality breaks down when little TRM is gained or when the NRM is almost fully demagnetized. This is when the remanence is bound by a physical restriction (e.g., remanence is restricted to exist only between ambient temperature and T_c), which produces an asymmetric distribution of remanence values, or when the remanence variance is dominated by sources of experimental noise that are non-Gaussian (e.g., $\delta\theta$, discussed below).

[42] The total variance of the TRM of Sample 1 is consistently higher than that of the NRM (Figure 2f). The variances of NRM and TRM that result from individual components of experimental noise are shown in Figures 2i and 2j, respectively. It should be noted that the KS test rejects the null hypothesis that some individual variance sources are normally distributed, which confirms that NRM and TRM variances cannot be described by Gaussian error propagation. For the sake of simplicity and first-order comparisons, however, individual variance sources are calculated assuming normality. Measurement errors ($\delta M_{(x,y,z)}$) dominate the NRM variance and are a major contributor to the TRM variance. When little TRM is gained the remanence variance that results from measurement reorientation noise, $\delta\theta$, is high. The KS test rejects the null hypothesis that remanence variance due to $\delta\theta$ is normally distributed, which is why the TRM is not normally distributed over this interval. As was noted above, the contribution from δBG is negligible. Similarly, all other sources of noise

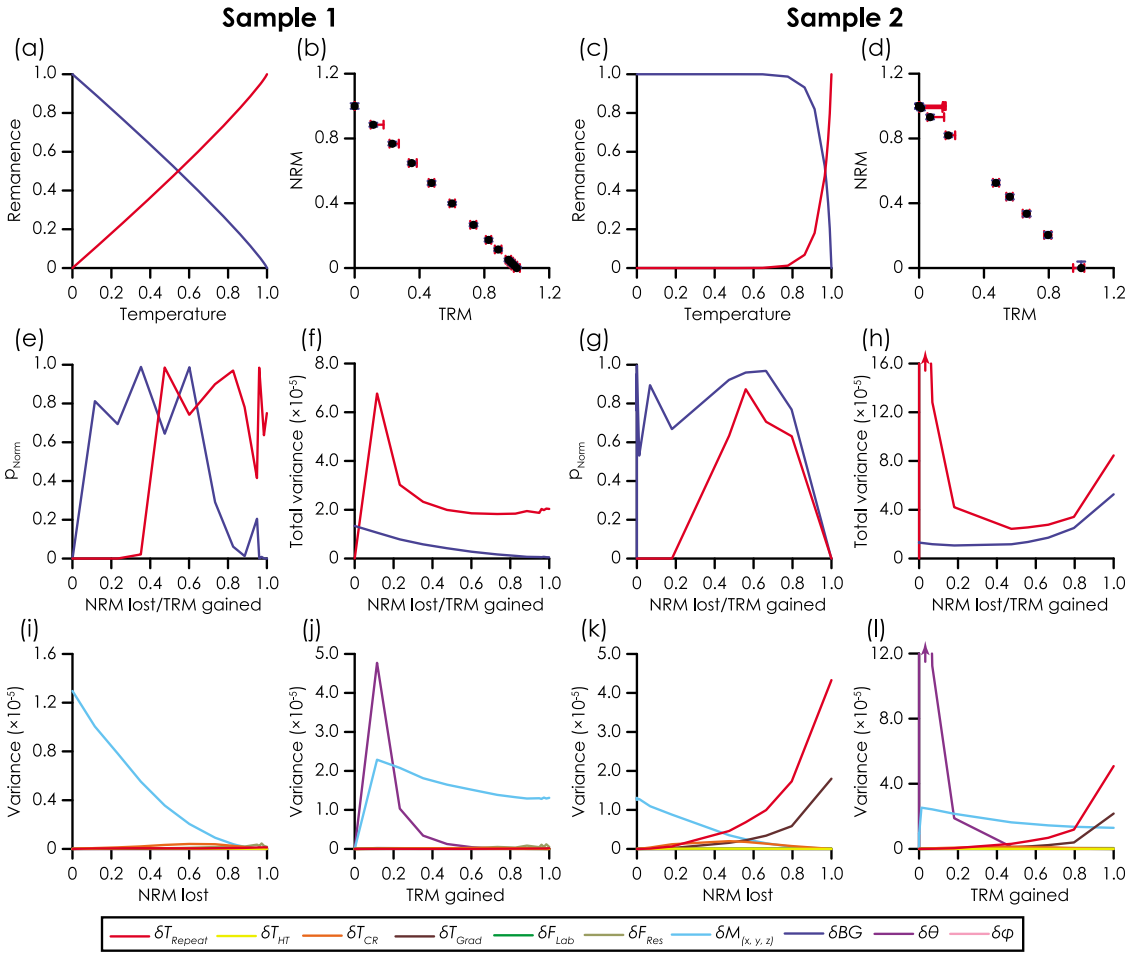

Figure 2. (a, c) The NRM demagnetization (blue line) and TRM acquisition (red line) with no experimental noise for Sample 1 and 2, respectively. (b, d) Arai plots from both samples. The black dots represent the ideal data and the error bars represent the maximum and minimum NRM (blue) and TRM values (red) obtained from the Monte Carlo simulations. (e, g) The probability that experimental noise produces normally distributed NRM (blue line) and TRM (red line) values for both samples. (f, h) The total (dimensionless) variance of NRM (blue line) and TRM (red line) for both samples. The contribution of variance from different sources for the (i) NRM and (j) TRM of Sample 1, and the (k) NRM and (l) TRM of Sample 2. In Figure 2h, TRM variance reaches a peak of $\sim 44 \times 10^{-5}$ when the TRM gained is $\leq 1.2\%$. This is due to variance from $\delta\theta$. For clarity, in both Figures 2h and 2l, TRM variance has been truncated.

make negligible contributions to the NRM and TRM variances of Sample 1.

[43] For Sample 2 the TRM variance is also consistently higher than the NRM variance (i.e., the scatter of points along the x-axis of the Arai plot is greater than along the y-axis; Figure 2h). The individual sources of variance indicate that δT_{Repeat} and δT_{Grad} contribute more to the NRM and TRM variances of Sample 2 than for Sample 1 (Figures 2k and 2l). This is intuitively expected given the narrow range of (un)blocking temperatures. The main contributions to the NRM variance are δT_{Repeat}, δT_{Grad}, and $\delta M_{(x,y,z)}$, with δT_{Repeat} dominating the total variance at high temperatures. The main sources of TRM variance are δT_{Repeat}, δT_{Grad}, $\delta M_{(x,y,z)}$, and $\delta\theta$. The TRM variance that results from reorientation noise ($\delta\theta$) is high at low temperatures when no NRM is demagnetized. For both samples, this large variance is the result of the vector subtraction of two near identical strong remanence vectors in the calculation of a relatively weak remanence vector (the TRM). When a sufficient amount of NRM is demagnetized and TRM acquired $\delta\theta$ drops to effectively zero.

4.2. The Effects on pTRM and pTRM tail checks

[44] The same simulations can be used to investigate the distributions of checks used to detect non-ideal behavior, as well as the contributions that the

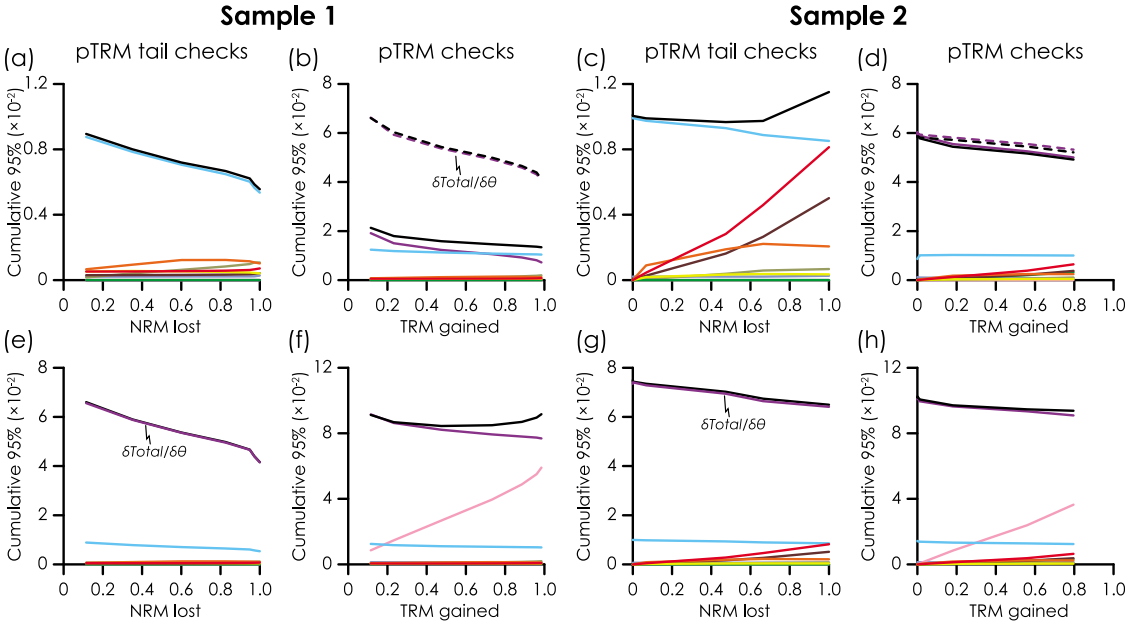

Figure 3. (a, b) The cumulative 95th percentile pTRM tail check and pTRM check values, respectively, for Sample 1. (c, d) The cumulative 95th percentile pTRM tail check and pTRM check values, respectively, for Sample 2. (e–h) The same as in Figures 3a–3d, but check values have been calculated by vector arithmetic. All check values are unnormalized (i.e., they are the remanence differences). The black lines represent the 95th percentile values from the models incorporating all source of experimental noise. The dashed lines in Figures 3b and 3d represent simulations where F_{Lab} was perpendicular to F_{Anc}. The remaining symbols are the same as in Figure 2. In Figures 3b, 3e and 3g the values resulting from all errors and those resulting from $\delta\theta$ coincide.

different sources of experimental noise make to these distributions. Check values are calculated as the absolute values of scalar differences and are unnormalized. No parametric distribution was found to adequately describe the distributions of the check values for either sample. Given that the check values are absolute deviations from zero we use the 95th percentile of the empirical cumulative distribution functions as a convenient nonparametric measure of the width of the distribution of check values. The parameter values that are used to select data are typically the maximum check values from all temperature steps below the highest temperature used for the best-fit on an Arai plot. Therefore, it is the combined distributions that control the maximum likely check value (i.e., the cumulated distribution of all previous check values). The cumulative 95th percentiles for pTRM and pTRM tail checks for both samples are shown in Figures 3a–3d. These values represent the 95th percentiles of the resultant distribution when all check value distributions from previous steps are combined. The cumulative 95th percentiles increase as high check values are added to the distribution and decrease as lower values are added.

[45] For both samples the check values are of a similar order of magnitude, but the values for Sample 2 tend to be higher. In both cases pTRM tail checks are dominated by measurement errors (Figures 3a and 3c). For both samples, residual fields, cooling rate, and hold time variations make noticeable contributions, but are not major sources of noise. δT_{Repeat} and δT_{Grad} are the main sources of temperature noise for pTRM tails checks for Sample 2.

[46] Considering pTRM checks (Figures 3b and 3d), for Sample 1 the main contributions are from measurement and orientation errors. For Sample 2, however, orientation errors are the largest contributor, with measurements errors the second largest. For both samples, however, when the applied field is perpendicular to the ancient field, measurement orientation errors dominate pTRM checks (dashed lines in Figures 3b and 3d). This increased influence of measurement orientation errors dramatically increases the total pTRM check discrepancy for Sample 1. In a situation where F_{Lab} is applied at an random angle with respect to F_{Anc} for a suite of samples, pTRM checks will tend to be controlled by measurement orientation errors. We note that the high TRM variance at low temperatures due to reorientation errors (e.g., Figures 2b and 2d) is the main reason for high pTRM checks. This implies that pTRM

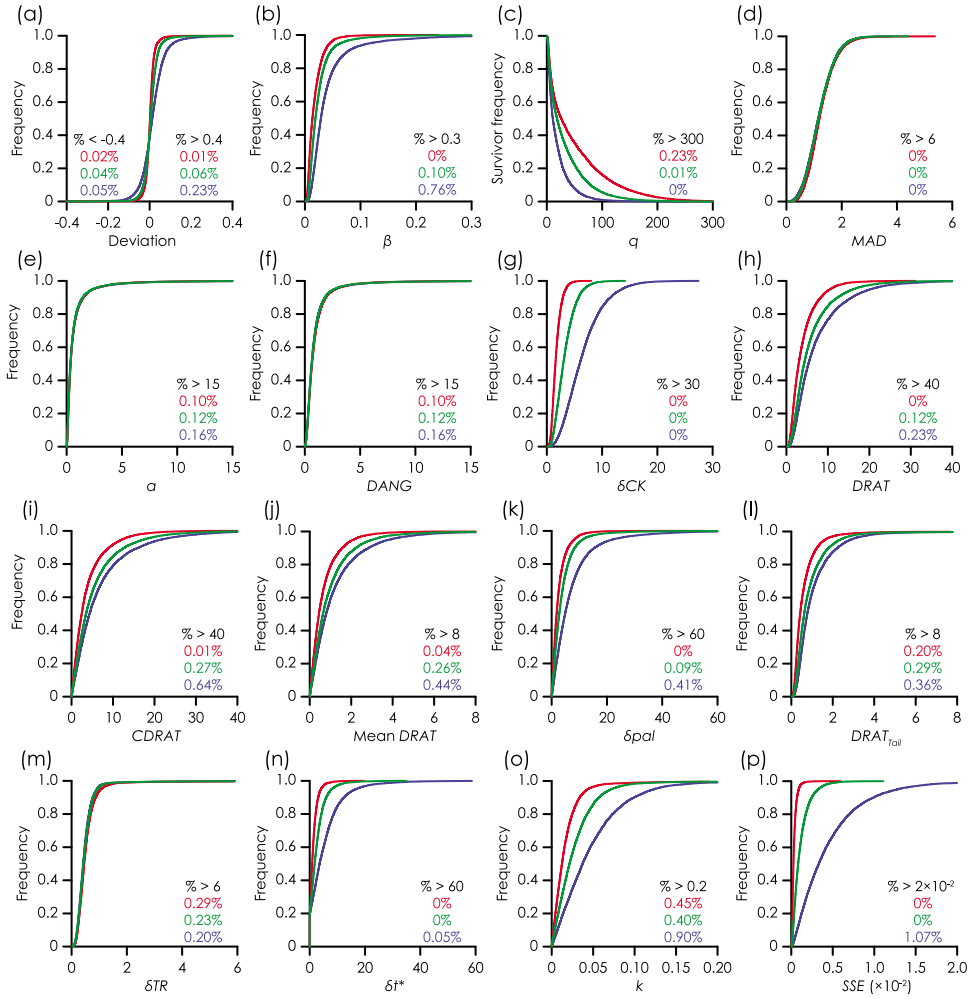

Figure 4. Empirical cumulative distribution functions (ECDFs) for various paleointensity parameters obtained from the Coe simulations with $f_{min} = 0.15$. The green lines represent the simulations where $F_{Lab} = F_{Anc}$, red lines where $F_{Lab} = 2F_{Anc}$, and blue lines where $F_{Lab} = \frac{1}{2}F_{Anc}$. For clarity some ECDFs have been truncated and the percentage of missing values is indicated on each plot. (a) Deviation of the paleointensity estimate from the expected value. (b–f) Arai plot and directional parameters. (g–k) pTRM check parameters. (l–n) pTRM tail check parameters. (o–p) Arai plot curvature parameters. Definitions of the parameters are given in the auxiliary material.

checks at low temperatures, even before normalization, are likely to be high in the absence of non-ideal behavior and should be treated with caution.

5. The Influence on Paleointensity Selection Parameters

[47] To investigate the effects that experimental noise has on the paleointensity parameters used to select data, additional Monte Carlo simulations of the Coe protocol experiment were performed. In these simulations, the parameters for the (un) blocking function are selected from the real data fits, but limited to the 93% (1,967 samples) best-fits to the real data (see the auxiliary material for further details). The procedure for the models is as follows.

[48] 1. Randomly select a (un)blocking spectrum from the real data fits.

[49] 2. Create a randomly oriented, idealized NRM.

[50] 3. Simulate the paleointensity experiment with F_{Lab} applied along the z-axis.

[51] 4. Randomly select a segment with a negative slope, comprising at least 4 points, and with a fraction (f) ≥ 0.15 and a gap factor (g) > 0.

[52] 5. Calculate the selection parameters for the best-fit segment.

[53] 6. Repeat steps 1–5 for 10^4 simulations.

Table 2. Descriptive Statistics of the Monte Carlo Simulations Along With Criteria Threshold Values Typically Used for Paleointensity Data Selection and the 95% Limits Determined From the Simulations[a]

Criterion	Typical Value	$F_{Lab} = \frac{1}{2}F_{Anc}$	$F_{Lab} = F_{Anc}$	$F_{Lab} = 2F_{Anc}$
		Coe Protocol		
Mean deviation	-	0.02	0.01	0.00
Scatter	-	6.9	4.3	3.3
Percent inaccurate	-	8.7	2.1	0.8
β	≤0.1–0.15	0.117	0.063	0.040
q	≥1–5	1.8	2.5	2.9
MAD	≤7–15	2.1	2.1	2.2
α	≤15	2.3	2.5	2.3
DANG	≤10–15	2.7	2.8	2.7
δCK	≤5–10	**13.5**	6.9	3.4
DRAT	≤7–10	**20.3**	**16.6**	**10.6**
CDRAT	≤10	**21.8**	**17.6**	**12.0**
Mean *DRAT*	≤3.5	**3.8**	3.0	2.0
δpal	≤5–10	**23.3**	**11.4**	7.7
$DRAT_{Tail}$	≤10	2.9	2.3	1.6
δTR	≤6.5–15	0.9	0.9	1.1
δt^*	≤3	**16.3**	**8.1**	**4.1**
k	≤0.164	0.122	0.069	0.042
SSE ($\times 10^{-2}$)	≤1.260	1.263	0.311	0.087
		Thellier-Thellier Protocol		
Mean deviation	-	0.00	−0.02	−0.05
Scatter	-	4.4	5.4	9.5
Percent inaccurate	-	3.4	11.1	24.0
β	≤0.1–0.15	0.077	0.127	**0.257**
q	≥1–5	1.5	**0.8**	**0.4**
MAD	≤7–15	7.8	**15.1**	**25.8**
α	≤15	3.9	7.0	13.7
DANG	≤10–15	7.5	15.1	**30.1**
δCK	≤5–10	6.7	3.4	1.7
DRAT	≤7–10	**10.2**	8.5	5.4
CDRAT	≤10	**11.1**	9.1	6.2
Mean *DRAT*	≤3.5	1.9	1.6	1.0
δpal	≤5–10	9.9	5.3	3.9
$DRAT_{Tail}$	≤10	8.9	**14.0**	**17.6**
δTR	≤6.5–15	2.4	4.9	10.8
δt^*	≤3	**14.0**	**7.0**	3.6
k	≤0.164	0.077	0.077	0.137
SSE ($\times 10^{-2}$)	≤1.260	0.295	0.194	0.705

[a]The minimum fraction for these simulations was $f_{min} = 0.15$. The criteria are defined in the auxiliary material. Bold font indicates situations when typically used values are likely to be too strict. Scatter is the standard deviation as a percentage of the mean.

[54] The criteria in step 4 are necessary to avoid unrealistic fits (e.g., fitting to a segment with $f = 0.01$).

[55] Three experiments were modeled, each with differing laboratory field strengths: $F_{Lab} = F_{Anc}$, $F_{Lab} = 2F_{Anc}$, and $F_{Lab} = \frac{1}{2}F_{Anc}$. Since F_{Lab} is directly used in the quantification of the magnitude of experimental noise (e.g., δF_{Lab}, δF_{Res}), F_{Anc} was varied to simulate different field ratios. The minimum acceptable fraction (f_{min}), used in step 4 above, was varied from 0.15 to 0.90. The results presented in this section are from a Coe protocol experiment including both pTRM and pTRM tail checks. For brevity the Aitken and IZZI protocols are not presented, but they yield near identical results and the below discussion is equally valid.

[56] Empirical cumulative distribution functions (ECDFs) for the deviation of the paleointensity estimates from the expected values and various paleointensity parameters commonly used to select data are shown in Figure 4 ($f_{min} = 0.15$). A table of quartiles for these distributions is given in the auxiliary material. The deviation is quantified as the logarithm of the intensity estimate normalized by the expected value. When the deviation is zero the estimate is exactly what is expected; positive and negative values represent over- and underestimates, respectively. Deviation values ≥−0.0953 and ≤0.0953 are accurate

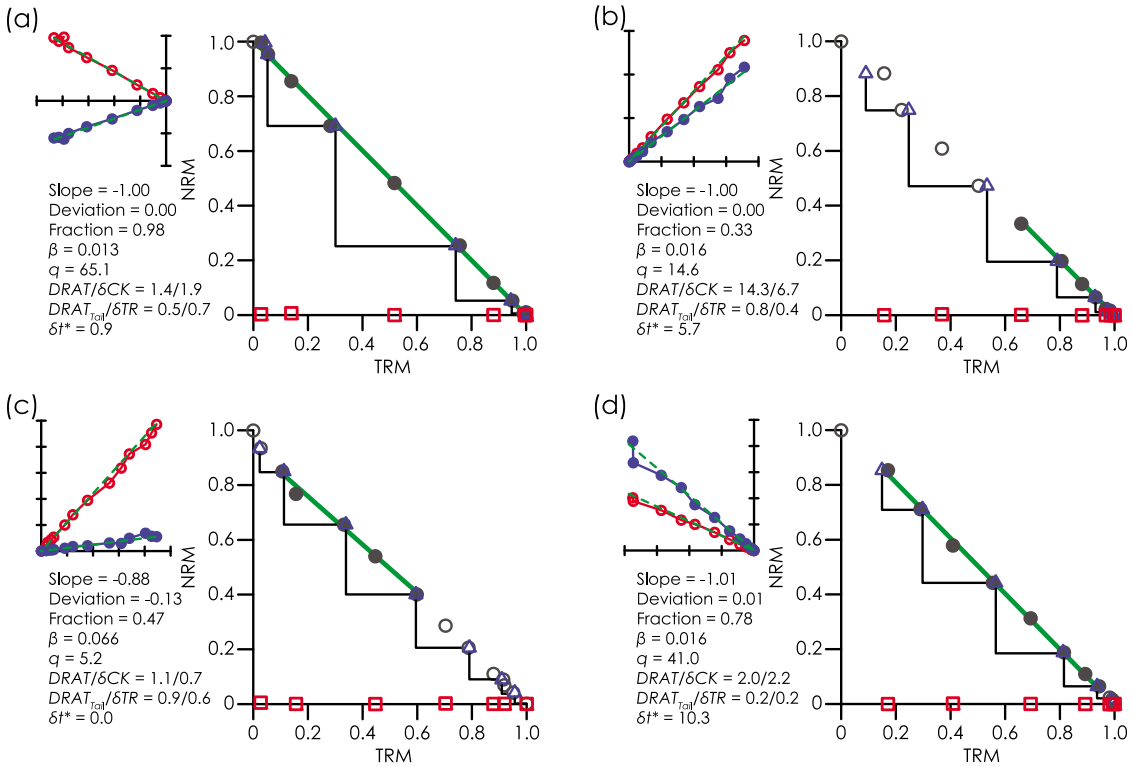

Figure 5. Example Arai and vector component diagrams from the simulations used to constrain the distributions of selection parameters. For all plots $F_{Lab} = F_{Anc}$. In the vector component diagrams, open (red) symbols represent the horizontal component and the closed (blue) symbols represent the vertical component. The dashed green lines represent the true direction of F_{Anc} (i.e., the NRM direction without noise). In the Arai plots the NRM and TRM have been normalized by their respective maxima. Solid circles represent the points used to calculate the best-fit linear segment (green line), blue triangles represent the pTRM checks, and red squares represent pTRM tail checks. (a) A near ideal sample that yields and accurate results. (b) Accurate, but has a high *DRAT* value due to a relatively low fraction. (c) A near ideal sample, but yields an inaccurate result. (d) Also a near ideal sample, but has a high δt^* value due to one noisy point that is identifiable in the vertical component of the NRM.

within a factor of 1.1 (i.e., accurate within ~10%) and are deemed to be accurate. Values outside of this ranged are classed as inaccurate. The definitions of the various selection parameters are given in the auxiliary material. Given that the fraction is largely controlled by the random selection of points from a uniform distribution it is near identical for all the three models and is not shown.

[57] A consistent feature of the different F_{Lab} models, is that when F_{Lab} is lower than F_{Anc} the results tend to be poorer (i.e., fewer accurate paleointensity estimates, more scattered data and with higher checks values). In the models where $F_{Lab} \geq F_{Anc}$, over 97.5% of the simulations yield accurate results, but when $F_{Lab} = \frac{1}{2}F_{Anc}$ only ~91% of the simulations yield accurate results (Table 2). The scatter (standard deviation) of the intensity estimates also has an applied field strength dependence (Table 2). The scatter increases from 3.3% when $F_{Lab} = 2F_{Anc}$, to 4.3% ($F_{Lab} = F_{Anc}$), and reaches a maximum of 6.9% when $F_{Lab} = \frac{1}{2}F_{Anc}$. All selection parameters exhibit a similar applied field dependence with the exception of directional parameters *MAD*, α, *DANG*, and the pTRM tail parameter δTR (Figure 4), which are all field invariant. It should be noted that F_{Anc} is varied in these models and these results are not related to changes in F_{Lab} directly influencing noise.

[58] Example Arai plots from the $F_{Lab} = F_{Anc}$ simulations are shown in Figure 5. The example in Figure 5a has well behaved data with low pTRM and tail check values and yields an accurate result. The example in Figure 5b yields an accurate result, but has a high *DRAT* value (14.3), which is likely to be rejected by typically used thresholds. As will be discussed below, this high value is related to the fraction of NRM used for the best-fit linear segment and a larger fraction would reduce *DRAT*; δCK would be unaffected. The fraction used for this fit ($f = 0.33$), however, would be accepted by many paleointensity studies. The example in Figure 5c is

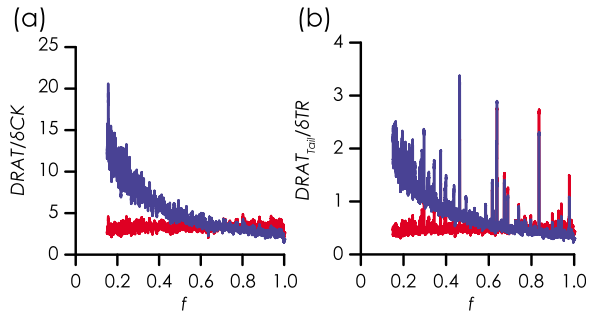

Figure 6. Fractional dependence of (a) pTRM checks DRAT (blue line) and δCK (red line) and (b) pTRM tail checks $DRAT_{Tail}$ (blue line) and δTR (red line). All values are from the $F_{Lab} = F_{Anc}$ Coe protocol simulation with $f_{min} = 0.15$ and have been smoothed using a 25 point running average.

a sample that would pass many typically used selection criteria (Table 2), but yields an inaccurate result. If the best-fit is extended to the next highest temperature step the result would be accurate (deviation = 0.094). The peak temperature of the original best-fit is 450°C, if alteration were to occur at high temperatures preventing the use of further steps the low-temperature segment is likely to be accepted and deemed to be a "reliable" result that passes selection. This highlights a seldom acknowledged issue with paleointensity data selection: even under the most ideal of conditions complete discrimination against inaccurate results may not be possible. In Figure 5d the Arai plot data are near ideal and yield a highly accurate results (deviation = 0.01), a high quality factor (41.0), and low check values. The δt^* value, however, is much higher than previously proposed cut-off values (Table 2). The high δt^* is due to one noisy NRM point, which randomly deviates toward the applied field direction (seen in the vector component diagram in Figure 5d). Although the pTRM tail check values are low, the correction for angular dependence used to calculate δt^* amplifies the noise. This sensitivity may reduce the ability of δt^* to detect non-ideal behavior.

[59] By design the *DRAT*-parameters, which normalize checks by the length of the best-fit line segment, have a fractional dependence to penalize checks based on small fractions. The fractional dependence of *DRAT*, δCK, $DRAT_{Tail}$, and δTR are shown in Figure 6. The data in this figure are from the $F_{Lab} = F_{Anc}$ model with $f_{min} = 0.15$ and have been smoothed using a 25 point running average. Data from the other field ratio models exhibit the same general trends, but with differing parameter values. For the *DRAT* pTRM check parameter, low fractions ($0.15 \leq f \lesssim 0.3$) produce high check values that, for most studies, would result in rejection. It should be emphasized that no non-ideal behavior, other than experimental noise, is present in these simulations and that if all inaccurate results (~2.1%) are removed this feature remains. This suggests that the fractional penalization of *DRAT* may be too strict. For pTRM tail checks experimental noise is unlikely to result in the rejection of data, but $DRAT_{Tail}$ still has a strong fractional dependence. Mean *DRAT* and *CDRAT* also exhibit a fractional dependence. The δ-parameters have no fractional dependence and only at fractions $\gtrsim 0.707$ (i.e., when the best-fit line is of equal length to the total NRM or TRM) do the average *DRAT*-parameter values fall below those of the δ-parameters.

5.1. Defining Parameter Limits Caused by Noise

[60] Limits of the parameter distributions can be defined by considering the distribution 95th percentiles. In the case of the quality factor (q [*Coe et al.*, 1978]), for which data are selected if they are above a critical value, we use the 5th percentiles. The limits obtained from the $f_{min} = 0.15$ simulations are given in Table 2. These values represent the limits of variability due to experimental noise and below these thresholds we cannot distinguish non-ideal behavior from the effects of noise. In practical terms, threshold values used for selection criteria should be less strict than these values otherwise near ideal samples that yield accurate results may be rejected.

[61] Most parameter 95% thresholds for $f_{min} = 0.15$ are below typical values used for data selection

Table 3. The Percentage of Results From Ideal Samples, Subject to Experimental Noise, That Are Rejected by Commonly Used Selection Criteria[a]

Criterion	Threshold	$F_{Lab} = \frac{1}{2}F_{Anc}$	$F_{Lab} = F_{Anc}$	$F_{Lab} = 2F_{Anc}$
		Coe Protocol		
δCK	>10	16.2	-	-
DRAT	>10	22.7	16.4	5.9
CDRAT	>10	21.4	15.9	7.4
Mean *DRAT*	≤3.5	6.2	-	-
δpal	>10	21.1	6.6	-
δt^*	>3	55.0	33.8	10.9
		Thellier-Thellier Protocol		
β	>0.15	-	-	19.3
q	<1	-	7.4	20.4
MAD	>15	-	5.1	22.5
DANG	>15	-	-	17.7
DRAT	>10	5.3	-	-
CDRAT	>10	6.4	-	-
$DRAT_{Tail}$	>10	-	12.2	18.2
δt^*	>3	51.6	28.6	8.1

[a]The minimum fraction for these simulations was $f_{min} = 0.15$.

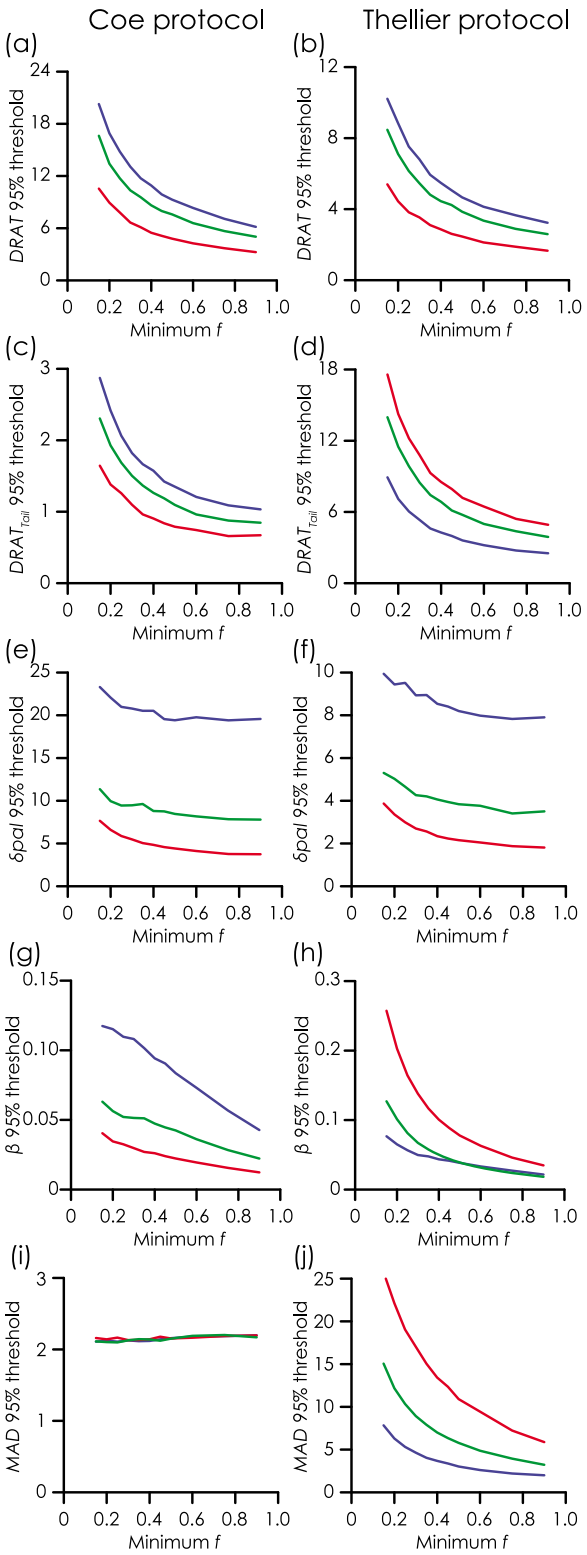

Figure 7. The dependence of selection parameter 95% thresholds on the minimum accepted fraction for the Coe and Thellier protocols. (a, b) *DRAT*, (c, d) $DRAT_{Tail}$, (e, f) δpal, (g, h) β, (i, j) *MAD*. The colors are the same as in Figure 4.

(Table 2). A number of parameters have 95% thresholds above typical selection criteria, notably *DRAT*, *CDRAT*, δpal, and δt^*. This means that these criteria are likely to reject hypothetical ideal samples that are subject only to experimental noise. Table 3 summarizes the percentage of results from these simulations that would be rejected by the typical selection criteria. The rejection rates range from 6.2% (i.e., from just above the 95% limit of detection, which is equivalent to a rejection rate of 5%) to a rejection rate of 55.0% (δt^*). The pTRM tail check parameter δt^* has the highest rejection rate of all criteria and this is a strong indication that this parameter is highly sensitive to noise and applied field strength.

[62] The fractional dependence of selection parameters discussed above also manifests as a fractional

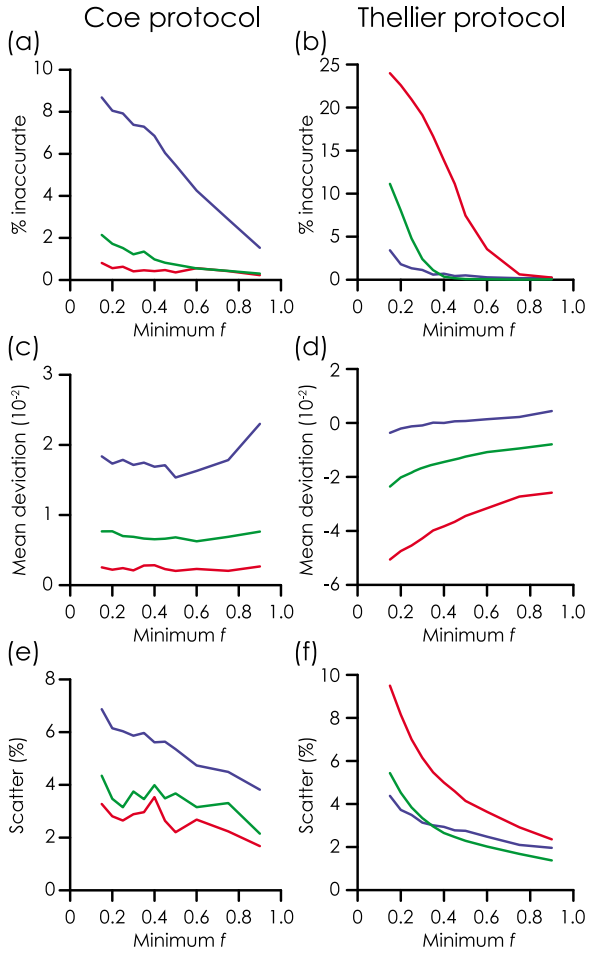

Figure 8. The dependence of descriptive statistics on the minimum accepted fraction for the Coe and Thellier protocols. (a, b) The percentage of inaccurate results. (c, d) The deviation of the mean results. (e, f) The scatter of the results as a percentage of the mean results. The colors are the same as in Figure 4.

dependence of the 95% thresholds. Some f_{min} dependent 95% parameter limits for the Coe simulations are shown in Figures 7a, 7c, 7e, 7g, and 7i and descriptive statistics in Figures 8a, 8c, and 8e. Additional parameters are shown in the auxiliary material. In general, as f_{min} increases the parameter 95% thresholds decrease, most notably the *DRAT*-parameters, which follow a power law decay with increasing f_{min}. The exception to this is q (shown in the auxiliary material), which increases due to its proportionality with f and $\frac{1}{\beta}$. The pTRM check δpal has only a weak dependence on f_{min}, which is most pronounced when $F_{Lab} > F_{anc}$ (Figure 7e). The percentages of inaccurate results when $F_{Lab} \geq F_{Anc}$ are consistently low ($\ll$5%), but an f_{min} of $\gtrsim$0.55 is needed to achieve a similar level for $F_{Lab} = \frac{1}{2}F_{Anc}$ (Figure 8a). The scatter of the results (Figure 8e) consistently falls below 5% for $f_{min} \geq 0.5$ and decreases to a minimum of $\sim$1.7–3.8% for $f_{min} = 0.9$.

[63] For all field strengths the deviation of the mean result is approximately constant (Figure 8c). Although the mean results are accurate, there is a small bias (up to $\sim$2.0%) towards overestimates of the true intensity. This bias is due to the asymmetric variance of the TRM at low and high temperatures. If we consider Figure 2d, experimental noise at low temperatures tends to produce apparently high TRM acquisition (the points can be easily shifted right on the plot). At high temperatures, where the extent of noise is largely controlled by T_c, noise tends to produce apparently low TRM acquisition (the points can be easily shifted left on the plot). The effect of this would be to produce a steeper slope, which would result in a small overestimate of the paleointensity. The small bias to high values is from best-fit slopes that use a large proportion of points from high or low temperatures.

6. Discussion

6.1. Other Models of Experimental Noise

[64] *Kono and Tanaka* [1984] proposed a method of estimating the variance of NRM and TRM on an Arai plot in order to find the most appropriate least-squares or maximum-likelihood estimator method for calculating the slope and associated error of the best-fit linear segment. Their method was based on Gaussian error propagation and, as tested in section 4.1, Gaussian error propagation is not appropriate for some individual error sources, most notably those associated with angular deviations. The NRM and TRM distributions, however, are found to be approximately Gaussian, so it may be possible that this approach is valid.

[65] We refer the reader to *Kono and Tanaka* [1984] for full details of their method, but we have updated their method as applied to a Coe protocol experiment (equation 15 in their paper) to use a beta distribution of (un)blocking temperatures and the error estimates constrained with real data. It is found that for both Sample 1 and Sample 2 the NRM variance is approximately one order of magnitude larger than calculated by our model and the TRM variance is about two orders of magnitude larger. Setting the angular uncertainties ($\delta\theta$ and $\delta\phi$) to zero reduces the variances to a comparable order of magnitude to our model. The presence of non-Gaussian noise is incompatible with the method of *Kono and Tanaka* [1984]. Our findings suggest that the validity of paleointensity estimates determined using the method of *Kono and Tanaka* [1984] are questionable and should be treated with caution. This highlights the importance of numerical methods to propagate uncertainties correctly.

6.2. The Generality of the Model

[66] Many of the noise sources incorporated into the model have been constrained by real data measurements that may not be general to every paleointensity study. For example, repeat temperature uncertainties are constrained from only the Natsuhara-Giken TDS-1 thermal demagnetizers at the Center for Advanced Marine Core Research and the measurement noise only from the 2G magnetometer at the IGGCAS. Other laboratories may use different equipment or the measured noise may have different statistical behavior. The effects of removing individual sources of experimental noise on the 95% thresholds was investigated and the tables for $f_{min} = 0.15$ and $f_{min} = 0.35$ are given in the auxiliary material. In general, when individual noise sources are removed from the model, the parameter 95% thresholds decrease by only a small amount. The exception is $\delta\theta$, which, if removed, reduces all threshold values. Therefore, $\delta\theta$ is the dominant source of noise and the other sources can vary considerably without affecting the overall results of these simulations. The generality of our estimate of reorientation uncertainty is difficult to ascertain. Although no other detailed measurements are currently available to estimate $\delta\theta$ for other users/laboratories, a first-order comparison can be made with the results of *Borradaile et al.* [2006] who investigated sample orientation using a Molspin spinner magnetometer. For a standard sized sample of diabase *Borradaile*

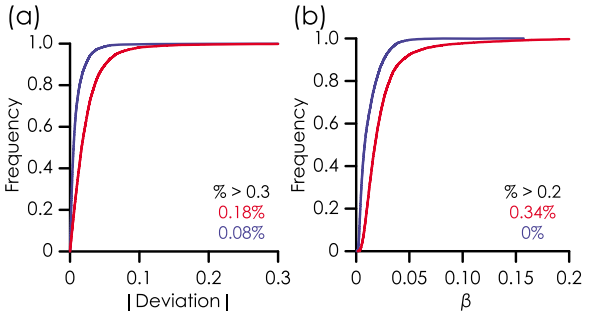

Figure 9. Comparison of the distributions of the (a) absolute deviation of the paleointensity estimates and (b) scatter of the best-fit line segments for a simulated Coe experiment ($F_{Lab} = F_{Anc}$ and $f_{min} = 0.15$) with all error sources included (red line) and with $\delta\theta$ and $\delta\phi$ set to zero (blue line).

et al. [2006] determined Fisher statistics of $\alpha_{95} = 0.5$ and $\kappa = 2723$ for 30 repeat measurements. For our 8 samples, the 25 reorientation measurements yield values of $\alpha_{95} = 0.65–0.82$ and $\kappa = 1260–2000$. These values are of a comparable magnitude and if the precision estimates remain the same and an additional five measurements were obtained the α_{95} values are likely to overlap. It should be noted that *Borradaile et al.* [2006] did not investigate the magnetometer measurement noise ($\delta M_{(x,y,z)}$) and that a single measurement using a Molspin magnetometer requires four separate sample reorientations. Despite these differences, this comparison suggests that our estimate of $\delta\theta$ is applicable to other studies and that our overall results are widely applicable.

6.3. Implications for Experiment Design

[67] Three main factors dominate the influence of experimental noise on paleointensity data, $\delta\theta$, $\delta M_{(x,y,z)}$, and δT_{Repeat}, which affect both paleointensity estimates and checks for non-ideal behavior. $\delta M_{(x,y,z)}$ and δT_{Repeat} are properties of the equipment and are largely out of the control of the user. The angular deviation, $\delta\theta$, which is the dominant influence on a paleointensity experiment, could be reduced by careful experimental design. Methods that fix a sample during the course of the entire paleointensity experiment (heating and measurement) will effectively eliminate $\delta\theta$ and $\delta\phi$. Such approaches, however, generally require specialized equipment such as the microwave method [*Hill and Shaw*, 1999] or paleointensity vibration sample magnetometers [e.g., *Le Goff and Gallet*, 2004]. Alternatively, specialized sample holders will reduce the influence of $\delta\theta$ [e.g., *Borradaile et al.*, 2006; *Böhnel et al.*, 2009]. This should be most effective if the sample is not removed from the holder for the duration of the paleointensity experiment (i.e., the sample is heated while within the holder). Holders that are used only for measurement will aid in the reduction of $\delta\theta$, but in some cases the reduction may be negligible. This may be particularly important for mini-samples (1 cm diameter), which are increasingly being used for paleointensity studies.

[68] In Figure 9 the distributions of paleointensity deviation and β from a Coe protocol model with all errors included are compared with a Coe model where both $\delta\theta$ and $\delta\phi$ are set to zero; in both models $F_{Lab} = F_{Anc}$ and $f_{min} = 0.15$. When angular deviations are eliminated the likelihood of obtaining an accurate result with a low scatter about the best-fit linear segment is increased. The reduction of β due to reduced angular deviation could potentially explain the reduction of β observed by [*Biggin*, 2010] when he compared the results of microwave paleointensity experiments to those obtained from thermal paleointensity experiments. A table of 95% threshold values for simulations with $\delta\theta$ and $\delta\phi$ set to zero is presented in the auxiliary material.

[69] A near consistent feature of the distribution of selection parameters is the tendency of producing poorer results when F_{Lab} is lower than F_{Anc} (Figure 4). That is to say there is an increased probability of less accurate results (Figure 4a), higher Arai plot scatter (from both β and SSE, Figures 4b and 4p, respectively), lower quality factors (Figure 4c), failure of both pTRM checks (Figures 4g–4k) and pTRM tails checks (Figures 4l–4o) is more likely, and Arai plots exhibit a higher degree of curvature (Figure 4o). These observations of paleointensity deviation and β are supported by the experimental data of *Morales et al.* [2006] who investigated the effects of applied field strength on natural samples. Their data also indicate that when F_{Lab} is $< F_{Anc}$ paleointensity estimates tend to be less accurate and Arai plots more scattered. *Tanaka and Kono* [1984] also investigated the effects of varying the strength of F_{Lab}. The results of *Tanaka and Kono* [1984] indicate that the scatter of the Arai plot is lowest when F_{Lab} is within a factor of ~2 of F_{Anc}. It should be noted that the paleointensity data of *Tanaka and Kono* [1984] used the variance analysis of *Kono and Tanaka* [1984] and, as noted above, this may yield inaccurate results. *Paterson et al.* [2010a] noted a high rate of data rejection from samples where F_{Lab} was ~5.6 times lower than F_{Anc}. Their interpretation was that the low F_{Lab} enhanced the

effects of MD grains, however, the results of our simulations suggest that experimental noise may have played a role in the failure of these samples. This may be the case for other studies.

[70] From our analysis of experimental noise it can be suggested that the most appropriate F_{Lab} should be $\geq F_{Anc}$. This, however, is based entirely on minimizing the effects of experimental noise acting on hypothetical ideal samples. A general approach of setting $F_{Lab} \gg F_{Anc}$ in real experiments would be inadvisable as this may exaggerate some types of non-ideal behavior (e.g., MD behavior [*Biggin*, 2006]).

6.4. Choice of Experimental Procedure

[71] Although the Coe protocol is the most widely used Thellier-type paleointensity protocol other protocols are used in modern studies [e.g., *Kissel et al.*, 2011; *Donadini et al.*, 2011; *Valet et al.*, 2010]. For an ideal sample, in the absence of noise, the Thellier, Coe, Aitken, and IZZI protocols will yield identical results. Additional simulations with $f_{\min} = 0.15$ indicate that the Coe, Aitken, IZZI, and the Coe protocol with no pTRM tail checks all yield the same 95% threshold values to within $\lesssim$0.5 percentage points (within the limits of the Monte Carlo approach). The exceptions to this are the directional parameters for the Aitken protocol with $F_{Lab} = 2F_{Anc}$, which are two times higher (*MAD* and α) or about three times higher (*DANG*) than the other protocols. Tables of the 95% threshold values for these simulations are given in the auxiliary material. In general, these four protocols all behavior in a similar fashion in the presence of experimental noise and the limits of detecting non-ideal behavior are the same. It should be noted, however, that the introduction of some degree of non-ideal behavior (e.g., alteration or grain size effects) will reduce the similarity between these protocols [e.g., *Biggin*, 2006].

[72] Considering the original Thellier-Thellier protocol, many of the 95% thresholds have an opposite field dependence to that seen for the Coe simulations (Table 2). For the Thellier-Thellier simulations when $F_{Lab} \geq F_{Anc}$ the results are more likely to be inaccurate, with a higher scatter around the best-fit line (Table 2). With the exception of pTRM checks and δt^*, all threshold values are higher when F_{Lab} is high. Unlike the other protocols, *MAD*, α, *DANG*, and δTR from the Thellier-Thellier simulations have a strong F_{Lab} dependence (Table 2). The pTRM checks follow the same F_{Lab} trend as the other protocols (i.e., higher values when $F_{Lab} < F_{Anc}$), but have 95% thresholds that are about half of those from the Coe simulations (Table 2). $DRAT_{Tail}$ has much higher 95% thresholds, which may be viewed as indicating the presence of non-ideal behavior, even though it is absent from these simulations. It is also noteworthy that when $F_{Lab} = F_{Anc}$ 11% of results are inaccurate and this increases to 24% when $F_{Lab} = 2F_{Anc}$ (Table 2). The higher pTRM tail checks, increased scatter and inaccuracy of results suggests that the original Thellier-Thellier protocol is more sensitive to experimental noise when compared with the Coe protocol. This is contrary to the findings of *Kono and Tanaka* [1984] who concluded that the errors in the Thellier-Thellier protocol were well balanced would lead to better performance.

[73] In general, for the Thellier-Thellier protocol, checks or parameters relating to TRM (i.e., pTRM checks) are lower than for Coe protocols, but those relating to NRM (i.e., Arai plot best-fit line, pTRM tail checks, and NRM directional parameters) tend to be much higher. The increased sensitivity of NRM related parameters to experimental noise is the result of the in-field heating steps and the vector arithmetic used to calculate the NRM. The in-field heating carries additional measurement and orientation noise (both field and measurement orientation), which are propagated through the vector sum used to calculate the NRM. The equivalent to Figure 2 for a Thellier-Thellier model is given in the auxiliary material. For both Sample 1 and 2, NRM variance is comparable to, or higher than TRM variance. This is particularly true at high temperatures where $\delta\theta$ and $\delta\phi$ dominate NRM variance. The difference between the Thellier and Coe protocols is not related to the small violation of Thellier's law of reciprocity (i.e., blocking and unblocking occur at the same temperature in an ideal sample) in the Coe protocol [*Dunlop and Özdemir*, 1997]. Full details are given in the auxiliary material, but it can be shown that if the maximum violation of reciprocity is assumed for the Coe protocol simulations, the only 95% thresholds affected are $DRAT_{Tail}$ and δTR, which should increase by $\lesssim$1 and $\lesssim$0.5 percentage points, respectively. It should be noted that this is an upper limit and the effect is likely to be smaller.

[74] The $f_{\min}$ dependence of the 95% thresholds and descriptive statistics for the Thellier-Thellier simulations are shown in Figures 7b, 7d, 7f, 7h, and 7j, and Figures 8b, 8d, and 8f. Although the *DRAT* 95% thresholds are lower than those from the Coe simulations, the $DRAT_{Tail}$ thresholds are much higher. An $f_{\min} \gtrsim 0.25$–0.35 is needed to reduce the $DRAT_{Tail}$ 95% thresholds to values that are below those typically used for data selection, but even at

Table 4. Descriptive Statistics of the Monte Carlo Simulations Along With Criteria Threshold Values Typically Used for Paleointensity Data Selection and the 95% Limits Determined From the Simulations of Different Protocols[a]

Criterion	Typical Value	Thellier	Coe	Coe No Tail Checks	Aitken	IZZI
Mean deviation	-	−0.02	0.01	0.01	0.01	0.01
Scatter	-	3.0	3.5	3.6	3.0	3.9
Percent inaccurate	-	1.1	1.4	1.1	1.1	1.1
β	≤0.1–0.15	0.058	0.051	0.050	0.050	0.050
q	≥1–5	4.7	6.5	6.9	6.9	6.8
MAD	≤7–15	7.9	2.1	2.1	2.3	2.2
α	≤15	3.7	2.0	1.9	2.2	2.0
DANG	≤10–15	7.6	2.3	2.1	3.2	2.3
δCK	≤5–10	3.4	6.8	6.7	6.9	6.9
DRAT	≤7–10	4.8	9.6	9.3	9.6	9.7
CDRAT	≤10	5.5	**10.9**	**10.5**	**10.9**	**10.7**
Mean *DRAT*	≤3.5	1.0	2.0	1.9	2.0	2.0
δpal	≤5–10	4.2	9.6	9.2	8.9	9.4
$DRAT_{Tail}$	≤10	7.4	1.4	-	-	1.4
δTR	≤6.5–15	5.0	1.0	-	-	0.9
δt^*	≤3	**7.1**	**8.4**	-	-	**8.1**
k	≤0.164	0.077	0.070	0.070	0.079	0.074
SSE ($\times 10^{-2}$)	≤1.260	0.196	0.319	0.311	0.322	0.320
		Percentage of Results Rejected by Typical Criteria				
CDRAT	>10	-	6.4	6.2	6.6	6.1
δt^*	>3	27.6	33.5	-	-	34.3

[a]For all simulations $f_{min} = 0.35$ and $F_{Lab} = F_{Anc}$. Bold font indicates situations when typically used values are likely to be too strict. Scatter is the standard deviation as a percentage of the mean.

$f_{min} = 0.9$ the thresholds are above the highest values from the Coe protocol (Table 2 and Figure 7c). For the scatter around the best-fit linear segment $f_{min} \gtrsim 0.4$ is needed to bring β below 0.1 (a typical selection criterion) for all values of F_{Lab} studies here (Figure 7h). The values of *MAD* for the Thellier-Thellier simulations have a strong F_{Lab} and fractional dependence (Figure 7j). This is in contrast to the Coe simulations in which *MAD* is F_{Lab} and fraction independent (Figure 7i). A typical selection criterion is to specify *MAD* ≤ 15; a fraction of $f_{min} \gtrsim 0.4$ will bring the *MAD* 95% thresholds below 15 for all Thellier-Thellier simulations.

[75] The percentage of inaccurate results from the $f_{min} = 0.15$ simulations is high, but drops rapidly with increasing f_{min} (Figure 8b). For $F_{Lab} \leq F_{Anc}$, $f_{min} \geq 0.35$–0.40 reduces inaccurate results to ≲1%, but $f_{min} \gtrsim 0.7$ is needed when $F_{Lab} = 2F_{Anc}$. The mean intensities are accurate, irrespective of f_{min} or F_{Lab}, but there is a small systematic underestimate, which is reduced as f_{min} increases (Figure 8d). As is the case for the Coe protocol simulations, the scatter of results from the Thellier-Thellier simulations decreases with increasing f_{min} (Figure 8f).

6.5. The Selection of Paleointensity Data

[76] From these simulations the majority of paleointensity estimates are accurate (within 10% of the expected value). The simulations with the lowest number of accurate estimates are those from the Thellier-Thellier protocol with $F_{Lab} = 2F_{Anc}$ where ∼24% of estimates are inaccurate. Most paleointensity studies aim to use an F_{Lab} value close to F_{Anc} and we therefore limit the following discussion to the $F_{Lab} = F_{Anc}$ simulations. In this case, the highest proportion of inaccurate results is 11%, again from the Thellier-Thellier protocol. The high β and *SSE* values, which are measures of Arai plot data scatter (Table 2) combined with the drop of inaccurate results with increasing f_{min}, suggests that scattered data on the Arai plot is the source of many inaccurate results from these simulations. The small proportion of inaccurate results compared with real paleointensity data (e.g., ∼55% of the historical data studied by *Paterson* [2011] yielded inaccurate estimates) and the lack of true non-ideal behavior in the simulations makes it difficult to define practical threshold values to use for data selection. As previously noted, however, the 95% thresholds in Table2 represent the upper confidence limit of parameter values that can be produced by experimental noise and below these values we cannot distinguish non-ideal behavior from noise.

[77] The strong fractional dependence of various selection parameters and descriptive statistics allows us to justify the definition of a minimum fraction for data selection. On the basis of reducing

the percentage of inaccurate results and reducing scatter we recommend a minimum fraction of 0.35 for all protocols. When $F_{Lab} \approx F_{Anc}$, this ensures low scatter of results (≤3.9%), low probability of inaccurate results (≤1.4%), and it lowers most of the 95% thresholds for fractional dependent parameters, which will increase their sensitivity to true non-ideal behavior. These values are summarized in Table 4.

[78] In general, most of the 95% thresholds of the modeled parameters (f_{min} = 0.35) are less than the critical values typically used to select data (Table 4), which implies that few (<5%) ideal samples subject to only noise are being rejected in real studies. This also means that some degree of non-ideal behavior is likely to be passing selection. To some extent this may be desirable, for example, a number of studies indicate that small pseudo-single domain sized grains, although in the strictest sense non-ideal, are capable of yielding accurate paleointensity estimates [e.g., *Shcherbakov and Shcherbakova*, 2001]. For the *CDRAT* and δt^* parameters, however, typically used thresholds are likely to be too strict. Such strict values result in the rejection of ∼6–34% of ideal samples, many of which yield accurate results (Table 4).

[79] Our analysis indicates that for the pTRM check parameters *DRAT*, *CDRAT*, and δpal the threshold values for data selection should be no less than ∼10. The fractional dependence of *DRAT* and *CDRAT* may result in these criteria being too strict if f is low. For example in the Coe protocol simulation with f_{min} = 0.15 and $F_{Lab} = F_{Anc}$ the *DRAT* threshold is 16.6, but this decreases to 9.6 when f_{min} = 0.35. The δCK pTRM check 95% threshold, however, has no fractional dependence. It should be noted that for the above mentioned three checks, at f_{min} = 0.35 the 95% thresholds are close to what are often viewed as more relaxed selection thresholds. If "stricter" criteria were to be used (e.g., $DRAT \leq 7$, or $\delta pal \leq 5$) it is likely that ideal data will be rejected.

[80] For the use of standard pTRM and pTRM tail checks, our simulations suggest that δ-parameters (i.e., δCK and δTR) are less sensitive to experimental noise and choice of best-fit line segment. The consistently lower 95% threshold values (Tables 2 and 4), the fractional independence (Figure 6), and the low applied field dependence of δTR (Figure 4i) support this argument. The fractional independence will make these parameters independent of the choice of best-fit linear segment, which may unknowingly suffer from user bias, and the reduced noise dependence should make these δ-parameters more sensitive to small degrees of non-ideal behavior. The efficacy of these selection parameters at excluding non-ideal behavior, however, needs to be tested further.

[81] The large 95% threshold values (compared with typical selection values) and the F_{Lab} dependence of δpal and δt^* are, in part, a result of how they are calculated: by vector arithmetic. The cumulative 95th percentile check values determined using vector arithmetic for the simulations in section 4.2 are shown in Figures 3e–3h. When vector arithmetic is used for pTRM tails checks $\delta\theta$ becomes a dominant source of noise, and for pTRM checks $\delta\phi$ makes significant noise contributions alongside $\delta\theta$. This does not necessarily mean that parameters based on vector arithmetic are inferior, but simply that the typical values seen from real data as well as their ranges and limits of detecting non-ideal behavior are different to what may be intuitively regarded as "suitable" for data selection.

[82] The typically used selection value for δt^* (≤3) is based on the phenomenological MD model of *Leonhardt et al.* [2004], which does not include experimental noise. This highlights the importance of incorporating experimental noise into paleointensity simulations if they are to be used to define data selection. The difference between this MD model and our SD model with noise suggests that, due to experimental noise, the ability for δt^* to identify non-ideal grain sizes may be reduced, but must be further investigated.

[83] In addition to defining the 95% threshold values, these simulations can also be used to define an upper limit for the deviation of an individual paleointensity estimate from the expected paleointensity. This can be achieved by taking the 95th percentile of the distribution of absolute deviations (e.g., the 95th percentile of the ECDFs shown in Figure 9a). This value represents the maximum likely degree of inaccuracy that results from experimental noise and cannot be avoided. For f_{min} = 0.35, the maximum degree of inaccuracy for all protocols using zero-field steps is ∼6% and is ∼7% for the Thelliier-Thellier protocol. This means that deviations of up to ∼6–7% could be caused by experimental noise and cannot be exclusively attributed to non-ideal behavior.

7. Conclusions

[84] Paleointensity data selection is a notoriously arbitrary process, but the models presented here

allow us to put the selection process on a more solid foundation. The approach outlined in this study allows us to investigate how various factors influence paleointensity selection and we come to the following conclusions.

[85] 1. By considering how experimental noise influences paleointensity data from hypothetical ideal samples it is possible to put a lower limit on our ability to detect non-ideal behavior. Paleointensity studies should not use selection threshold stricter than these limits for risk of excluding near ideal samples that are subject only to experimental noise.

[86] 2. For experiments using zero-field steps these limits are universal, but the behavior of the original Thellier-Thellier method is sufficiently different to require different limits. A set of selection criteria defined for a Coe experiment should not be used for a Thellier experiment and *vice versa.*

[87] 3. It is possible for ideal samples, subject to expected levels of experimental noise, to yield inaccurate results that cannot be discriminated by data selection.

[88] 4. Reorientation uncertainty during measurements is the dominant noise source that can affect paleointensity data and should the main priority for noise reduction for all paleointensity studies. Methods that fix the sample for the duration of the experiment or that use specialized sample holders should reduce this influence.

[89] 5. The choice of laboratory field can greatly influence the effects of experimental noise and we recommend using a field strength close to that of the ancient field.

[90] 6. When selecting data careful consideration must be given to the interplay of different parameters, specifically how the choice of fraction influences other criteria. The sensitivity of *DRAT*-parameters to fraction may lead to the rejection of well behaved data. The δ-parameters (δCK and δTR), however, have no fractional dependence and yield consistent results irrespective of the choice of best-fit linear segment and may be more suitable for consistent data selection.

[91] 7. How the choice of minimum fraction influences selection parameters and the statistics describing accuracy and scatter allows us to justify the use of a minimum fraction for data selection. We strongly recommend specifying $f \geq 0.35$ for all experimental protocols. This will reduce the likelihood of accepting inaccurate results that are caused by experimental noise. It also lowers many of the 95% thresholds. The lowering of these thresholds, below which we cannot distinguish non-ideal behavior from experimental noise, serves to increase the sensitivity of these parameters to non-ideal effects.

[92] 8. In the presence of experimental noise unavoidable inaccuracies of up to ~6–7% should be expected when $f \geq 0.35$. Any bias from non-ideal factors will be in addition to this baseline degree of inaccuracy. Studies that distinguish between accurate and inaccurate results should use deviations >6–7% to distinguish between the two groups.

[93] 9. The sensitivity of some parameters to noise and applied field may diminish their ability to discriminate against non-ideal factors. This is most notable for pTRM checks *DRAT*, *CDRAT* and δpal, and pTRM tail check δt^*. The typically used threshold values for these parameters should be relaxed or alternative parameters used. The sensitivity of δt^* to noise from a single point (e.g., Figure 5d) suggests that δt^* may not a robust parameter, but the efficacy of δt^* to distinguish non-ideal factors need to be tested further.

[94] 10. Future models, particularly those aimed at defining the selection of paleointensity data, must incorporate experimental noise in order to provide a sufficient degree of realism to have practical applications. Work is currently under way to incorporate experimental noise into the phenomenological MD model of *Biggin* [2006].

Acknowledgments

[95] Shuhui Cai, Baochun Huang, Xinlin Ji, Chengying Liu and Huafeng Qin are all thanked for providing data. Some thermal demagnetization data were obtained from the MagIC Paleomagnetic database (http://earthref.org/MAGIC/) and we are grateful for the efforts of the MagIC team. We thank Karl Fabian, Lisa Tauxe, and an anonymous Associate Editor for their thorough reviews. Lisa Tauxe is further thanked for providing additional data. We also thank David Dunlop for his comments. G.A.P. acknowledges funding from a Young International Scientist Fellowship from the Chinese Academy of Sciences (CAS; grant 2009Y2BZ5), Natural Science Foundation China (NSFC) grant 41050110132, and CAS grant KZCX2-YW-Q08 held by Y.P. A.J.B. acknowledges funding from a Natural Environment Research Council (NERC) Advanced Fellowship (NE/F015208/1). Y.Y. acknowledges funding from Japan Society for the Promotion of Science (JSPS) KAKENHI (23740340).

References

Aitken, M. J., A. L. Allsop, G. D. Bussell, and M. B. Winter (1988), Determination of the intensity of the Earth's

magnetic field during archaeological times: Reliability of the Thellier technique, *Rev. Geophys.*, *26*, 3–12, doi:10.1029/RG026i001p00003.

Biggin, A. J. (2006), First-order symmetry of weak-field partial thermoremanence in multi-domain (MD) ferromagnetic grains: 2. Implications for Thellier-type palaeointensity determination, *Earth Planet. Sci. Lett.*, *245*, 454–470, doi:10.1016/j.epsl.2006.02.034.

Biggin, A. J. (2010), Are systematic differences between thermal and microwave Thellier-type palaeointensity estimates a consequence of multidomain bias in the thermal results?, *Phys. Earth Planet. Inter.*, *180*, 16–40, doi:10.1016/j.pepi.2010.03.005.

Biggin, A. J., M. Perrin, and M. J. Dekkers (2007), A reliable absolute palaeointensity determination obtained from a non-ideal recorder, *Earth Planet. Sci. Lett.*, *257*, 545–563, doi:10.1016/j.epsl.2007.03.017.

Böhnel, H. N., D. Michalk, N. Nowaczyk, and G. G. Naranjo (2009), The use of mini-samples in palaeomagnetism, *Geophys. J. Int.*, *179*, 35–42, doi:10.1111/j.1365-246X.2009.04260.x.

Borradaile, G. J., B. S. Almqvist, and K. Lucas (2006), Specimen size and improved precision with the Molspin spinner magnetometer, *Earth Planet. Sci. Lett.*, *241*, 381–386, doi:10.1016/j.epsl.2005.11.050.

Coe, R. S. (1967), Paleo-intensities of the Earth's magnetic field determined from Tertiary and Quaternary rocks, *J. Geophys. Res.*, *72*, 3247–3262, doi:10.1029/JZ072i012p03247.

Coe, R. S., S. Grommé, and E. A. Mankinen (1978), Geomagnetic paleointensities from radiocarbon-dated lava flows on Hawaii and the question of the Pacific nondipole low, *J. Geophys. Res.*, *83*, 1740–1756, doi:10.1029/JB083iB04p01740.

Dodson, M. H., and E. McClelland-Brown (1980), Magnetic blocking temperatures of single-domain grains during slow cooling, *J. Geophys. Res.*, *85*, 2625–2637.

Donadini, F., S.-Å. Elming, L. Tauxe, and U. Hålenius (2011), Paleointensity determination on a 1.786 Ga old gabbro from Hoting, central Sweden, *Earth Planet. Sci. Lett.*, *309*, 234–248, doi:10.1016/j.epsl.2011.07.005.

Draeger, U., M. Prévot, T. Poidras, and J. Riisager (2006), Single-domain chemical, thermochemical and thermal remanences in a basaltic rock, *Geophys. J. Int.*, *166*, 12–32, doi:10.1111/j.1365-246X.2006.02862.x.

Dunlop, D. J., and Ö. Özdemir (1997), *Rock Magnetism: Fundamentals and Frontiers*, *Cambridge Stud. in Magn.*, vol. 3, Cambridge Univ. Press, New York.

Fabian, K. (2001), A theoretical treatment of paleointensity determination experiments on rocks containing pseudo-single or multi domain magnetic particles, *Earth Planet. Sci. Lett.*, *188*, 45–58, doi:10.1016/S0012-821X(01)00313-2.

Fabian, K. (2009), Thermochemical remanence acquisition in single-domain particle ensembles: A case for possible overestimation of the geomagnetic paleointensity, *Geochem. Geophys. Geosyst.*, *10*, Q06Z03, doi:10.1029/2009GC002420.

Halgedahl, S. L., R. Day, and M. D. Fuller (1980), The effect of cooling rate on the intensity of weak-field TRM in single-domain magnetite, *J. Geophys. Res.*, *85*, 3690–3698.

Hill, M. J., and J. Shaw (1999), Palaeointensity results for historic lavas from Mt Etna using microwave demagnetization/remagnetization in a modified Thellier-type experiment, *Geophys. J. Int.*, *139*, 583–590, doi:10.1046/j.1365-246x.1999.00980.x.

Holtzman, W. H. (1950), The unbiased estimate of the population variance and standard deviation, *Am. J. Psychol.*, *63*, 615–617.

Huang, B., J. A. Piper, Y. Wang, H. Y. He, and R. Zhu (2005), Paleomagnetic and geochronological constraints on the post-collisional northward convergence of the southwest Tian Shan, NW China, *Tectonophysics*, *409*, 107–124, doi:10.1016/j.tecto.2005.08.018.

Huang, B., J. D. A. Piper, H. He, C. Zhang, and R. Zhu (2006), Paleomagnetic and geochronological study of the Halaqiaola basalts, southern margin of the Altai Mountains, northern Xinjiang: Constraints on neotectonic convergent patterns north of Tibet, *J. Geophys. Res.*, *111*, B01101, doi:10.1029/2005JB003890.

Huang, B., J. D. A. Piper, C. Zhang, Z. Li, and R. Zhu (2007), Paleomagnetism of Cretaceous rocks in the Jiaodong Peninsula, eastern China: Insight into block rotations and neotectonic deformation in eastern Asia, *J. Geophys. Res.*, *112*, B03106, doi:10.1029/2006JB004462.

Incropera, F. P., D. P. DeWitt, T. L. Bergman, and A. S. Lavine (2007), *Fundamentals of Heat and Mass Transfer*, 6th ed., John Wiley, Hoboken, N. J.

Kissel, C., H. Guillou, C. Laj, J. C. Carracedo, S. Nomade, F. Perez-Torrado, and C. Wandres (2011), The Mono Lake excursion recorded in phonolitic lavas from Tenerife (Canary Islands): Paleomagnetic analyses and coupled K/Ar and Ar/Ar dating, *Phys. Earth Planet. Inter.*, *187*, 232–244, doi:10.1016/j.pepi.2011.04.014.

Kono, M., and H. Tanaka (1984), Analysis of the Thelliers' method of paleointensity determination: 1. Estimation of statistical errors, *J. Geomagn. Geoelectr.*, *36*, 267–284.

Krása, D., C. Heunemann, R. Leonhardt, and N. Petersen (2003), Experimental procedure to detect multidomain remanence during Thellier-Thellier experiments, *Phys. Chem. Earth*, *28*, 681–687, doi:10.1016/S1474-7065(03)00122-0.

Le Goff, M., and Y. Gallet (2004), A new three-axis vibrating sample magnetometer for continuous high-temperature magnetization measurements: Applications to paleo- and archeointensity determinations, *Earth Planet. Sci. Lett.*, *229*, 31–43, doi:10.1016/j.epsl.2004.10.025.

Leonhardt, R., D. Krása, and R. S. Coe (2004), Multidomain behavior during Thellier paleointensity experiments: A phenomenological model, *Phys. Earth Planet. Inter.*, *147*, 127–140, doi:10.1016/j.pepi.2004.01.009.

Liu, C. Y., and R. X. Zhu (2009), Geodynamic significances of the Emeishan Basalts, *Earth Sci. Frontiers*, *16*, 52–69, doi:10.1016/S1872-5791(08)60082-2.

Morales, J., A. Goguitchaichvili, L. M. Alva-Valdivia, and J. Urrutia-Fucugauchi (2006), Further details on the applicability of Thellier paleointensity method: The effect of magnitude of laboratory field, *C. R. Geosci.*, *338*, 507–513, doi:10.1016/j.crte.2006.02.007.

Muxworthy, A. R., X. L. Ji, V. Ridley, Y. Pan, L. Chang, L. Wang, and A. P. Roberts (2011), Multi-protocol palaeointensity determination from middle Brunhes Chron volcanics, Datong Volcanic Province, China, *Phys. Earth Planet. Inter.*, *187*, 188–198, doi:10.1016/j.pepi.2011.06.005.

Nagata, T., Y. Arai, and K. Momose (1963), Secular variation of the geomagnetic total force during the last 5,000 years, *J. Geophys. Res.*, *68*, 5277–5281.

Néel, L. (1949), Théorie du traînage magnétique des ferromagnétiques en grains fins avec applications aux terres cuites, *Ann. Géophys.*, *5*, 99–136.

Pan, Y., M. J. Hill, and R. Zhu (2005), Paleomagnetic and paleointensity study of an Oligocene-Miocene lava sequence from the Hannuoba Basalts in northern China, *Phys. Earth Planet. Inter.*, *151*, 21–35, doi:10.1016/j.pepi.2004.12.004.

Paterson, G. A. (2009), Assessment of the usefulness of lithic clasts from pyroclastic deposits as paleomagnetic recorders, PhD, thesis, Univ. of Southampton, Southampton, U. K.

Paterson, G. A. (2011), A simple test for the presence of multidomain behaviour during paleointensity experiments, *J. Geophys. Res.*, *116*, B10104, doi:10.1029/2011JB008369.

Paterson, G. A., A. R. Muxworthy, A. P. Roberts, and C. Mac Niocaill (2010a), Assessment of the usefulness of lithic clasts from pyroclastic deposits for paleointensity determination, *J. Geophys. Res.*, *115*, B03104, doi:10.1029/2009JB006475.

Paterson, G. A., A. P. Roberts, A. R. Muxworthy, C. Mac Niocaill, L. Gurioli, J. G. Viramonté, C. Navarro, and S. Weider (2010b), Paleomagnetic determination of emplacement temperatures of pyroclastic deposits: An under-utilized tool, *Bull. Volcanol.*, *72*, 309–330, doi:10.1007/s00445-009-0324-4.

Pullaiah, G., E. Irving, K. L. Buchan, and D. J. Dunlop (1975), Magnetization changes caused by burial and uplift, *Earth Planet. Sci. Lett.*, *28*, 133–143, doi:10.1016/0012-821X(75)90221-6.

Qin, H. F., H. Y. He, Q. Liu, and S. Cai (2011), Palaeointensity just at the onset of the Cretaceous normal superchron, *Phys. Earth Planet. Inter.*, *187*, 199–211, doi:10.1016/j.pepi.2011.05.009.

Riisager, P., and J. Riisager (2001), Detecting multidomain magnetic grains in Thellier palaeointensity experiments, *Phys. Earth Planet. Inter.*, *125*, 111–117, doi:10.1016/S0031-9201(01)00236-9.

Shcherbakov, V. P., and V. V. Shcherbakova (2001), On the suitability of the Thellier method of palaeointensity determinations on pseudo-single-domain and multidomain grains, *Geophys. J. Int.*, *146*, 20–30, doi:10.1046/j.0956-540x.2001.01421.x.

Shcherbakova, V. V., N. V. Lubnina, V. P. Shcherbakov, S. Mertanen, G. V. Zhidkov, T. I. Vasilieva, and V. A. Tsel'movich (2008), Palaeointensity and palaeodirectional studies of early Riphaean dyke complexes in the Lake Ladoga region (northwestern Russia), *Geophys. J. Int.*, *175*, 433–448, doi:10.1111/j.1365-246X.2008.03859.x.

Tanaka, H., and M. Kono (1984), Analysis of the Thelliers' method of paleointensity determination: 2. Applicability to high and low magnetic fields, *J. Geomagn. Geoelectr.*, *36*, 285–297.

Tauxe, L. (2010), *Essentials of Paleomagnetism*, Univ. of California Press, Berkeley.

Tauxe, L., and K. P. Kodama (2009), Paleosecular variation models for ancient times: Clues from Keweenawan lava flows, *Phys. Earth Planet. Inter.*, *177*, 31–45, doi:10.1016/j.pepi.2009.07.006.

Tauxe, L., P. Gans, and E. A. Mankinen (2004a), Paleomagnetism and $^{40}Ar/^{39}Ar$ ages from volcanics extruded during the Matuyama and Brunhes Chrons near McMurdo Sound, Antarctica, *Geochem. Geophys. Geosyst.*, *5*, Q06H12, doi:10.1029/2003GC000656.

Tauxe, L., C. Luskin, P. Selkin, P. Gans, and A. Calvert (2004b), Paleomagnetic results from the Snake River Plain: Contribution to the time-averaged field global database, *Geochem. Geophys. Geosyst.*, *5*, Q08H13, doi:10.1029/2003GC000661.

Thellier, E. (1938), Sur l'aimantation des terres cuites et ses applications géophysiques, *Ann. Inst. Phys. Globe Univ. Paris*, *16*, 157–302.

Thellier, E., and O. Thellier (1959), Sur l'intensité du champ magnétique terrestre dans le passé historique et géologique, *Ann. Géophys.*, *15*, 285–376.

Valet, J.-P., E. Herrero-Bervera, J. Carlut, and D. Kondopoulou (2010), A selective procedure for absolute paleointensity in lava flows, *Geophys. Res. Lett.*, *37*, L16308, doi:10.1029/2010GL044100.

Walton, D. (1979), Geomagnetic intensity in Athens between 2000 BC and AD 400, *Nature*, *277*, 643–644, doi:10.1038/277643a0.

Yamamoto, Y., and H. Tsunakawa (2005), Geomagnetic field intensity during the last 5 Myr: LTD-DHT Shaw palaeointensities from volcanic rocks of the Society Islands, French Polynesia, *Geophys. J. Int.*, *162*, 79–114, doi:10.1111/j.1365-246X.2005.02651.x.

Yamamoto, Y., H. Tsunakawa, and H. Shibuya (2003), Palaeointensity study of the Hawaiian 1960 lava: Implications for possible causes of erroneously high intensities, *Geophys. J. Int.*, *153*, 263–276, doi:10.1046/j.1365-246X.2003.01909.x.

Yu, Y. J., and D. J. Dunlop (2002), Multivectorial paleointensity determination from the Cordova Gabbro, southern Ontario, *Earth Planet. Sci. Lett.*, *203*, 983–998, doi:10.1016/S0012-821X(02)00900-7.

Yu, Y. J., and L. Tauxe (2006), Effect of multi-cycle heat treatment and pre-history dependence on partial thermoremanence (pTRM) and pTRM tails, *Phys. Earth Planet. Inter.*, *157*, 196–207, doi:10.1016/j.pepi.2006.04.006.

Yu, Y. J., L. Tauxe, and A. Genevey (2004), Toward an optimal geomagnetic field intensity determination technique, *Geochem. Geophys. Geosyst.*, *5*, Q02H07, doi:10.1029/2003GC000630.

Zheng, Z., X. Zhao, and C.-S. Horng (2010), A new high-precision furnace for paleomagnetic and paleointensity studies: Minimizing magnetic noise generated by heater currents inside traditional thermal demagnetizers, *Geochem. Geophys. Geosyst.*, *11*, Q04Y08, doi:10.1029/2010GC003100.

Zhu, R. X., Y. X. Pan, H. Y. He, H. F. Qin, and S. M. Ren (2008), Palaeomagnetism and $^{40}Ar/^{39}Ar$ age from a Cretaceous volcanic sequence, Inner Mongolia, China: Implications for the field variation during the Cretaceous normal superchron, *Phys. Earth Planet. Inter.*, *169*, 59–75, doi:10.1016/j.pepi.2008.07.025.